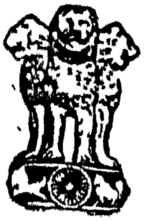


सत्यमेव जयते

INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI

I.A.R.I.6.
GIP NLK—H-3 I.A.R.I.—10-5 55—15,000

PROCEEDINGS

OF THE

ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL 197

LONDON
Published for the Royal Society by the
Cambridge University Press
Bentley House, N.W.1

7 July 1949

Printed in Great Britain at the University Press, Cambridge
(Brooke Crutchley, University Printer)
and published by the Cambridge University Press
Cambridge, and Bentley House, London

Agents for Canada and India: Macmillan

CONTENTS

SERIES A VOLUME 197

No. A 1048. 11 May 1949

PAGE

The presentation of scientific information. By E. N. da C. Andrade, F.R.S. (Plate 1) 1

On the bond lengths and interatomic distances in certain molecules and crystals. By W. Hume-Rothery, F.R.S. 17

Theory of drafting of wool slivers. I. By D. R. Cox 28

Explosion waves and shock waves. VII. The velocity of detonation in cast T.N.T. By W. B. Cybulski, (the late) W. Payman and D. W. Woodhead. (Plates 2 to 4) 51

On the self-energies and cross-sections of orthodox quantum mechanics. By H. S. Green 73

The structure of the reaction zone in a flame. By S. F. Boys and J. Corner . 90

Starting potentials of electrodeless discharges. By E. W. B. Gill and A. von Engel 107

Momentum and energy diffusion in the turbulent wake of a cylinder. By A. A. Townsend 124

No. A 1049. 7 June 1949

The catalyzed dehydration of acetaldehyde hydrate, and the effect of structure on the velocity of protolytic reactions. By R. P. Bell, F.R.S., and W. C. E. Higginson 141

Some remarks on integral equations with kernels: $L(\xi_1 - x_1, \ldots, \xi_n - x_n; \alpha)$. By G. Kreisel 160

The catalytic decomposition of simple heterocyclic compounds. I. Preparation and properties of molybdenum sulphide and nickel sulphide catalysts. By E. H. M. Badger, R. H. Griffith and W. B. S. Newling 184

The catalytic decomposition of simple heterocyclic compounds. II. Reaction kinetics and mechanism. By R. H. Griffith, J. D. F. Marsh and W. B. S. Newling 194

The hydrodynamical theory of film lubrication. By W. F. Cope 201

Crystal growth from solution. By S. P. F. Humphreys-Owen. (Plates 5 and 6) . 218

Flexural vibrations of the walls of thin cylindrical shells having freely supported ends. By R. N. Arnold and G. B. Warburton 238

On the quadratic form $x^2 - 7y^2$. By P. Varnavides 256

Charge distribution and energy levels of trapped electrons in ionic solids. By J. H. Simpson 269

No. A 1050. 22 June 1949

PAGE

Spontaneous deformation of protocatechuic acid crystals. By R. W. Wood, For.Mem.R.S. (Plates 7 to 12) 283

The oxidation of copper and the reactions of hydrogen and carbon monoxide with copper oxide. By W. E. Garner, F.R.S., T. J. Gray and F. S. Stone. (Plate 13) 294

The application of semi-conductivity measurements in the study of catalysis on copper oxide. By T. J. Gray 314

The constitution of the aluminium-rich aluminium-cobalt-copper alloys, with special reference to the role of transitional metals in alloy formation. By P. C. L. Pfeil and G. V. Raynor. (Plate 14) 321

The refinement of atomic parameters by the technique known in X-ray crystallography as 'the method of steepest descents'. By A. D. Booth 336

Studies in polymerization. III. The polymerization of methyl methacrylate. By C. H. Bamford and M. J. S. Dewar 356

Studies in polymerization. IV. The velocity constants in the polymerization of *p*-methoxystyrene. By D. W. E. Axford 374

Evaporation and surface structure of liquids. By G. Wyllie 383

Dislocations as a cause of mechanical damping in metals. By J. D. Eshelby . 396

The determination of static and dynamic yield stresses using a steel ball. By R. M. Davies 416

No. A 1051. 7 July 1949

The magnetic field within the earth. By E. C. Bullard, F.R.S. 433

Microscopy by reconstructed wave-fronts. By D. Gabor. (Plates 15 to 17) . . 454

The thermal equilibrium at the tropopause and the temperature of the lower stratosphere. By R. M. Goody 487

The scattering of high-energy neutrons by protons and the law of interaction between nucleons. By E. H. S. Burhop and H. N. Yadav 505

Recent advances in the technique of submarine gravity surveying. By R. I. B. Cooper 523

The passage of a perfect fluid through a critical cross-section or 'throat'. By A. M. Binnie 545

Residual lattice strains in plastically deformed polycrystalline metal aggregates. By G. B. Greenough 556

Electroviscosity. III. Sedimentation phenomena in ionic liquids. G. A. H. Elton 568

Index 573

The presentation of scientific information

By E. N. da C. Andrade, F.R.S.

(*Lecture delivered before the Royal Society Scientific Information Conference on* 23 *June* 1948)

[Plate 1]

When the organizers of this conference did me the honour to suggest that I should give this lecture on some general aspect of the theme that is engaging our deliberations, I was somewhat at a loss to know what plan to adopt. Like most of those present I have, in one way and another and at various times, been engaged in the communication of scientific information, but that does not necessarily make it an easy subject to discuss. Not to weary you with my doubts and inward debates, I finally decided to attempt a brief historical survey of the subject, for to see how one's difficulties arose is sometimes a step towards their solution. In any case I hope that the considerations that I shall lay before you may prove a mild and gentle, if not soporific, interlude in the strenuous deliberations upon which you are engaged.

I wish, then, to consider how the communication of scientific information—more particularly the communication of original scientific discovery, of even the smallest degree of importance and interest—has grown to be the very complex matter that it is to-day. The methods of communication that we shall have to consider are, I suppose, the spoken word, the book and the scientific periodical, which to-day includes both the periodical publishing original papers and the periodical publishing abstracts of papers. You may be surprised that I have included the spoken word, or, if you prefer it, personal communication, but I think that, if you consider, you will come to the conclusion that this personal method still retains its importance. I do not include radio under this head, because we are, I think, concerned rather with a first record of novel scientific results than with the popular exposition of general scientific advances. I would remind you that even in our times when students from all over the world flock to such a laboratory as Rutherford's (I choose a man no longer with us to avoid a selection among the living that might be invidious) it is not so much to avail themselves of material instruments—the instruments used by Rutherford could be bought by laboratories in any country—nor to hear a statement of what the leader has already achieved, which can be read in the scientific journals, as to hear from the leader's mouth what is going on in his head at the moment, to learn of scientific progress in the making. The writing of a paper comes, in general, after the results have already been communicated to a small circle in the laboratory and their friends. The importance of the spoken word has not yet vanished, and it would do no harm if some of our great men, whose audiences are attracted by their fame and retained in their seats by considerations of courtesy, were to remember that he who uses the spoken word should pay some attention to acquiring proficiency in the use of the instrument, just as he takes trouble in learning to handle any other instrument of his profession. Occasionally the performance brings to mind the man who was asked if he could play the violin and replied that he did not know—he had never tried. The ability to present scientific facts clearly in a lecture is not a gift from heaven; I doubt very much if anybody gives a really good first lecture. At any rate, taking one of the best British lecturers of modern times, one who was particularly good at communicating new scientific advances, Sir William Bragg, it is recorded of him that in his early days as professor he was one of the least impressive lecturers, but by careful application he attained that high eminence in the art for which he was later known.

Let us now look back to the beginnings of science and consider—very briefly—the communication of knowledge in the early days of Greek science. The spoken word at that time was supreme, although, of course, the written scroll transmitted the teaching. It would weary you were I to attempt to picture the great schools of antiquity—Plato's Academy, the Schools of Athens, of Pergamon, of Crotona are familiar to all of you. What we are apt to forget, however, is the freedom with which men travelled great distances in those days in order to learn by the spoken word of the advances in science and in philosophy. Plato's Academy was at Athens, but he thought nothing of visiting Sicily, and the students and philosophers of classical times travelled from Greece to Asia Minor, the Black Sea, Alexandria, Italy and North Africa as a matter of course. Since the days of the pre-eminence of Athens there has probably never been a time when personal intercourse among the learned of Europe was more set about with difficulties than it is to-day, which enhances for us the importance of the written word.

I would not, of course, wish to minimize the part played by the manuscript in early science. The great historian Mommsen, for instance, says, writing of the times of Julius Caesar: 'Reading had become a fashion, nay a mania.... The superior officer was seen in the camp-tent with the obscene Greek romance, the statesman in the senate with the philosophic treatise in his hands.' The two libraries of Alexandria, finally destroyed by the Saracens, must, in their prime, have contained at least half a million books, of which a fair proportion dealt with the science of the ancient world, and there were considerable libraries at Rome, Athens, Ephesus, Timgad, Constantinople and other centres of culture. Mention of Alexandria calls to mind that the Greeks had no sacred books, which was at that time advantageous for scientific advance, as illustrated by the traditional reply of the Caliph to the general who wished to spare the Alexandrian books from destruction—'if these writings of the Greeks agree with the Book of God, they are useless and need not be preserved: if they disagree they are pernicious, and ought to be destroyed.' The word of the teacher was transmitted by document—we have Tennyson's picture of Lucretius, passing

> To turn and ponder those three hundred scrolls
> Left by the Teacher, whom he held divine

—a considerable library. We have, then, in classical times, the great teacher, attracting scholars from all over the learned world, and the manuscript document as the two important methods of scientific communication. There were, needless to say, no scientific periodicals.

Turning to the Middle Ages and the revival of learning, we have the same methods for the communication of scientific ideas, the manuscript and the spoken word, with the same very great importance attached to the spoken word, partly on account of the cost of books and partly on account of the enthusiasm which brought students from all over Europe to centres of learning. In the fifteenth century there were some sixty universities in Europe; even when it is admitted that some of them were of very minor importance the number is impressive. The travels undertaken by learned men, no matter what their particular interests, and the international character of learning in what are called 'the Dark Ages', were extraordinary. Relations among men of science of all countries were, of course, much facilitated by the existence of what we no longer have, an international language of learning. Latin was the medium of expression in university and medical school, practically all books were written in Latin, and it must be remembered that it was not until round about 1700 that the man of science regularly expressed himself in his native language. Newton's *Principia* (1687) was written in Latin, his *Opticks* (1704) in English, and while, of course, I know that many scientific books were written in English, French and Italian before 1700, and many in Latin after 1700, yet it is a convenient approximate date to take for the transition.

And here, perhaps, I may interpolate a word as to the language most used in science to-day. In 1934 Sir Charles Sherrington made an examination of the *World List of Scientific Periodicals* and counted the number of scientific periodicals published in five great languages. The result was:

Russian	1,833	scientific	periodicals
Italian	1,667	,,	,,
German	6,186	,,	,,
French	5,013	,,	,,
English	13,494	,,	,,

The total number for the other four languages does not much exceed the number for English. It is quite clear, then, that if there is an international language in science to-day it is English. I take pleasure in quoting the last sentence of Sherrington's letter,* which runs as follows: 'The results cited above make evident the special responsibility resting on those contributors to science whose language is English to spare no pains to write it worthily of the great role entrusted to it as a medium of scientific thought to-day.'

Let me return to the Middle Ages and consider a few examples of the international character of learning. Saint Thomas Aquinas was born near Naples round about 1225 and, after a preliminary education at Monte Cassino and Naples, later studied in Paris for three years and then went to Cologne. He returned to Paris to teach there, and later taught at Rome, Paris again and Naples. He died on his way to Lyons. You will say that he was not a scientist, but he took an interest in the science of his times and was a very great scholar. Perhaps you prefer to consider Roger Bacon, who was born in Somerset and studied at Oxford (though not at Brasenose, as tradition asserts, for it was not founded until after his death). He taught for years at Paris, which was the centre of the intellectual life of the Middle Ages. Perhaps a later name, a forerunner of modern science, will appeal more, Vesalius, a Fleming, who studied at Louvain, Montpellier and Paris, and then went to Italy. There in 1537, at the age of twenty-two, he was appointed professor at Padua, where he spent seven years. Later he went to Basle, where he lectured while his great book, *De Humani Corporis Fabrica*, was passing through the press there. After that he went to Spain, Cyprus, Jerusalem, and died on his way to Padua, whither he had been summoned to succeed Fallopius. Later, our John Dee, who was born in 1527 and wrote a very remarkable preface to the first English translation of Euclid, as well as valuable books on navigation, travelled as a young man in the Low Countries, forming friendships with many mathematicians of fame, studied at Louvain, a great centre of learning where van Helmont was later, and went to Paris, where he lectured to audiences so large that the hall would not hold them, with the result that many of the students clambered up outside the windows, so that they could see him even if they could not hear him. He corresponded with men in the Universities of Orleans, Cologne, Heidelberg, Strasbourg, Verona, Padua, Ferrara, Bologna, Urbino and Rome. His subsequent travels took him to Antwerp and the Low Countries again, Bremen, Lubeck, Cracow, Prague and Central Europe in general. Cardan, the celebrated Italian doctor and mathematician, travelled over the whole of Italy and in Germany, Switzerland, the Tyrol, France, England, Scotland and the Low Countries. Certainly personal contacts and private correspondence were among the main means of communicating knowledge in those times.

---

* *Nature*, **134**, 624 (1934).

But I must not weary you, who are filled with the glory of the present and the faith of the future, with further stories of the remote past. The coming of printing lent a much increased importance to the written word as a means of communication. It made the words of the Arabian and classical writers on science widely accessible—and it must be remembered that medieval medicine was dominated by the Arabs. The first Euclid was printed in 1482 at Venice; the works of Pliny, Hippocrates, Galen, Aristotle, Celsus and Dioscorides were printed frequently in the period between 1469 and the end of the century.

The first great scientific revelations made in print were in the *De Revolutionibus* of Copernicus in 1543, and in Vesalius's *De Humani Corporis Fabrica* which appeared in the same year; the seventeenth century opened with Gilbert's *De Magnete*; Harvey's *De Motu Cordis* appeared in 1627; Kepler's *Astronomia Nova* and *Harmonices Mundi* in 1609 and 1619; Galileo's *Dialogo* in 1632. By 1660, the year when the Royal Society was founded, the printed book was thoroughly established as a means of communication, although the private letter was still favoured—see, for instance, the *Commercium Epistolicum* printed at Oxford in 1658 in which letters on mathematics exchanged between Brouncker, Kenelm Digby, Fermat, Frenicle, Wallis and Schooten are printed. As examples of earlier writers of letters that conveyed and discussed the scientific news of the day may be mentioned Descartes and Mersenne—the latter using 'barbarous Latin and pitiful French', according to Duhem.

The foundation of the Royal Society marks the beginning of a new stage in scientific communication, the scientific periodical, as well as the inception of the modern scientific society. Actually the Accademia del Cimento was founded earlier, in 1657, and published in 1666 an account of its experiments in a book which exercised a great influence in its time, the *Saggi di Naturali Esperienze*, which, appearing in Italian, the vernacular (an unusual thing), was translated into Latin and English, possibly into other languages. The Accademia lived, however, for only ten years. The Accademia dei Lincei, which lasted from 1600 to 1630, was even earlier, and had the support of Galileo. But of the scientific societies which have endured the Royal Society was the first, and its *Transactions*, which have now been a channel of scientific communication for 283 years, were first published in 1665. In the beginning these *Transactions* were published, not by the Society, but by the Secretary on his own account, and contained descriptions not only of what was communicated to the Society by Fellows, but also of work done abroad—thus in the very first number there is 'An Accompt of the Improvement of Optick Glasses at Rome', 'A Catalogue of the Philosophical Books publisht by Monsieur de Fermat at Tholouse, lately dead', and so on. The *Transactions* had, in fact, something of an international character, for foreigners sometimes carried on scientific controversy in its pages. When the *Transactions* temporarily ceased publication in 1676 Robert Hooke personally undertook the publication called *Philosophical Collections*, which went on until 1682. These *Collections* contain, besides communications from Fellows, accounts of matters of scientific interest abstracted from foreign journals or books, such as the description of Lana's flying chariot taken from his *Prodromo*, with a criticism by Hooke, and the review of Borelli's *De Motu Animalium*. In 1683 the *Transactions* appeared once again, and since that time have continued without interruption. The scientific journal was thus established.

Actually the first number of the *Journal des Sçavans*, a privately published periodical, appeared a few months before the first number of the *Transactions*. The *Journal* published accounts of scientific discussions; reviews of books, classical, literary and theological as well as scientific; and notices of men of science newly dead, such as the notice of Fermat which appeared in an early number. The *Journal* was, perhaps, more closely akin to *Nature* than to the *Transactions*. It must have been popular, for it appeared in two formats, as a folio* at Paris and as a 12mo in Amsterdam. It is amusing to note, as showing how little reliance can be placed on some learned authorities, that Martha Ornstein† quotes Zedler's 'flüchtigblütige Franzosen haben Ekel vor Folianten' ('volatile Frenchmen dislike folios') to explain why the *Journal des Sçavans* was so small and the *Acta Eruditorum* ('huge folios' according to her, but in fact ordinary quartos) so large. Actually the standard edition of the *Journal* is slightly larger than the *Acta*.

The *Acta Eruditorum* first came out in 1682 and was published in Latin at Leipzig, with the support of the Elector of Saxony. A large part of it is taken up with book reviews, not only of scientific, but also of theological, philosophical, legal and other books. It also contains, however, original contributions by outstanding men of science. The volume for 1688, for instance, includes a 7-page review of Stillingfleet's *The Antiquities of the British Churches*, a 6-page review of Boyle's *Disquisition about Final Causes* and a 12-page review of Newton's *Principia*, and Latin translations of papers that had appeared in our *Philosophical Transactions*, as well as original papers by Papin, and Jacob Bernoulli. The *Acta* thus gave a general account of what was going on in the world of learning. Another German scientific journal which appeared about this time was the *Miscellanea naturae curiosorum*, while the famous Thomas Bartholinus brought out in Copenhagen the *Acta medica et philosophica hafniensia*, which lasted from 1673 to 1680. The *Nouvelles de la republique des lettres* appeared in 1684, in Holland, and gave accounts of affairs in the scientific world. It did not last for long, but had imitators. Thus by the end of the seventeenth century scientific journals, both those designed for the publication of original work and those of a more popular nature, had made their appearance and established their places as organs of communication between men of science.

---

* While having the usual form of a quarto (page 8¾ × 6¾ in.) the book collates as a folio.

† *The Role of Scientific Societies in the Seventeenth Century*, 1928.

Perhaps I may here refer to a curious undertaking of which I can find no mention except a mere citation in certain catalogues, the *Weekly Memorials for the Ingenious*. These '*Memorials*' were weekly collections of reviews of current books, mostly of scientific interest 'with some other curious *Novelties* relating to *Arts* and *Sciences*'. Fifty parts appeared for the year 1682, and were afterwards bound in one volume.* In the preface the author (— Beaumont, according to the only authority† that mentions him) refers to Oldenburg and his *Philosophical Transactions* and to Hooke and his *Philosophical Collections* and says of the Royal Society: 'If the R.S. shall think my endeavours in this kind any way subservient to their *designes*, it may animate my *industry* to perform things in the best manner I may, none being more devotedly their Servant than my self.' Among the authors of whose books account is given are Caspar Bartholinus, Borelli, Jacob Bernoulli, Boyle, Fermat, Malpighi and Papin. Many of the reviews are taken from the *Journal des Sçavans*, a few from the *Acta Eruditorum*. Such a weekly review of scientific books has no counterpart, I believe, to-day.

Before leaving the period I should, perhaps, mention the works of an encyclopaedic character that became popular in the sixteenth and seventeenth centuries, and did much to spread knowledge. Conrad Gesner, of Zürich, was a naturalist who brought out a great *Historia Animalium* in four volumes (1551–58), from which our Topsel compiled his *Histories of Foure-Footed Beastes* (1607). Aldrovandus brought out a great book on birds, Moufet on insects. Everyone knows of Gerarde's *Herball* (1597), founded on the earlier work of Dodoens, and typical of a number of herbals which have been fully described by Agnes Arber.‡ All these books were beautifully illustrated. Reference may also be made to the appearance of German works which gave a general account of all that was known in wide fields of physical science at the time, including the most recent advances—forerunners, in a way, of the modern

---

* The history of this periodical is somewhat complicated. There were two completely different and rival issues for a certain period.

† Halkett and Laing, *Dictionary of Anonymous and Pseudonymous English Authors*.

‡ *Herbals: Their Origin and Evolution*, 1912; 2nd ed. 1938. Cambridge University Press.

German 'Handbuch'. Such works were Schwenter's *Deliciae Physico-Mathematicae* (1651) in three parts of some 600 pages each; Gaspar Schott's *Magia Universalis* (1658) in four volumes, and his *Mechanica Hydraulicopneumatica* (1657), in which the first account of Guericke's pump was given;* and Sturm's *Collegium Experimentale* (1676). All these works were intended for a wide circle of readers, and if, in general, they contained little in the way of first statement of original discoveries, did much to spread scientific knowledge.

At the opening of the eighteenth century we find a publication answering a new need. *Miscellanea Curiosa* in three volumes (1705–1707) contained a selection of the most important papers which 'being register'd in the Voluminous Journals of the *Royal Society* are amongst a multitude of less useful Matters, so Obscurely hid, that but very few inquisitive Gentlemen ever so much as heard of them'. The names of Halley, Newton, Clopton Havers, John Wallis and de Moivre occur freely as authors, Halley being particularly well represented. The papers are not abridged, for, as the preface says, they are in many cases already abridgments themselves. What we have here, then, is in some way a forerunner of the 'Progress Reports', which are an essential feature of our modern machinery of information. We also find, in 1704, the issue of the first technical dictionary in any language, the *Lexicon Technicum* of John Harris, of which a second volume appeared in 1710, notable for containing the first appearance of Isaac Newton's note *De Natura Acidorum*, to the importance of which Academician Sergei Vavilov recently directed attention.† The *Lexicon Technicum*, which was in English, contained articles on, for instance, Newcomen's newly invented steam-engine, and on the microscope, illustrated with a plate of John Marshall's 'new invented double microscope', and went into several editions (5th ed. 1736). We can scarcely mention encyclopaedias as a medium for the communication of scientific knowledge without passing reference to the great French *Encyclopédie* of Diderot, which was originally to have been a French translation of Ephraim Chambers's *Cyclopaedia*, but developed into a very much larger work, a vast encyclopaedia of science and, in particular, applied science and manufacture. D'Alembert was responsible for the mathematical articles, and Voltaire was among the contributors. The first volume appeared in 1751, the last volume of the letterpress in 1765 and the final volume of the plates in 1772. It caused much political controversy: why, it is hard to understand nowadays.

The first half of the eighteenth century saw the foundation of the Academy of Sciences of St Petersburg in 1725, now the Academy of Sciences of the U.S.S.R. Just after the end of the war the 220th anniversary of its foundation was made the occasion of a magnificent celebration, which some of us had the good fortune to attend. The first volume of its *Mémoires* appeared in 1728. Not long after this the *Proceedings of the Royal Swedish Academy* began (1739), to be followed in a few years by the *Proceedings of the Royal Danish Academy* (1745). At that time, of course, Germany was a body of independent States of which Hanover was the first to produce a scientific journal in the *Verhandlungen der Königlichen Akademie der*

---

* Guericke's own description first appeared in his *Experimenta nova, ut vocant, Magdeburgica de vacuo spatio*, in 1672.

† *Royal Society Newton Tercentenary Celebrations*, 1947.

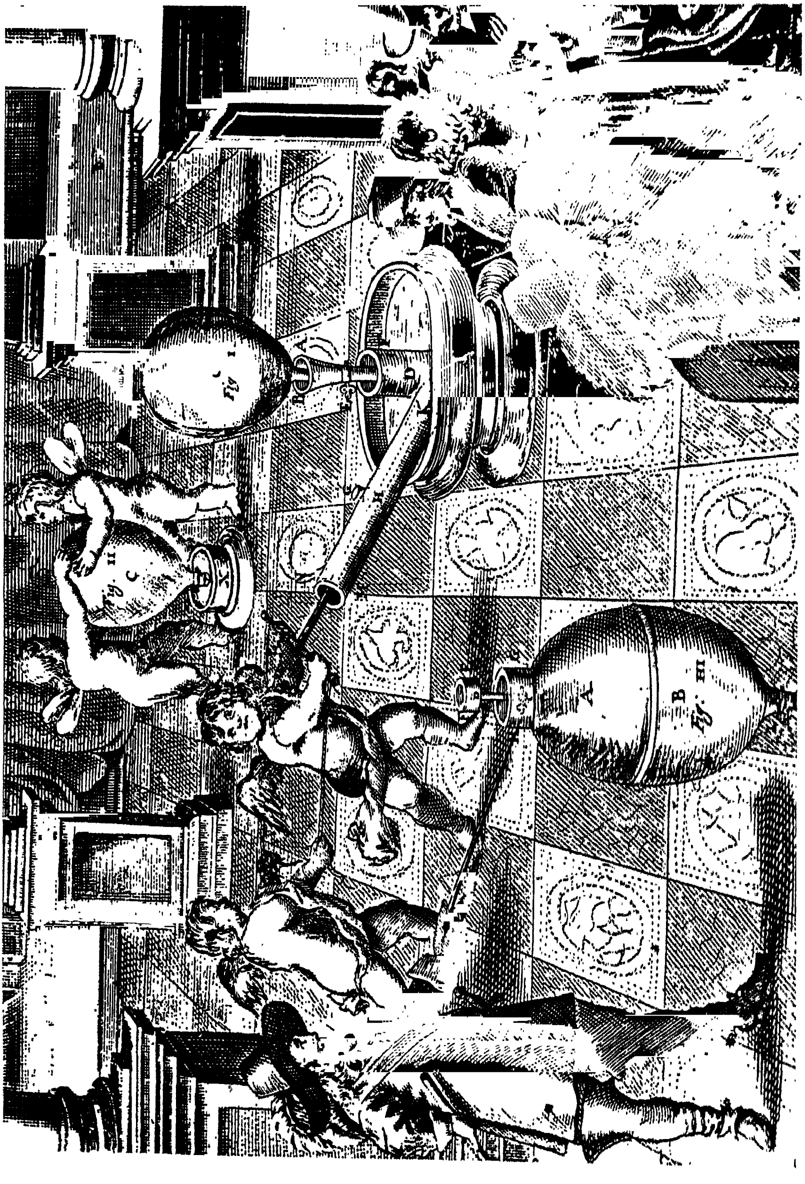


FIGURE 1. The first representation of Guericke's air pump, from Gaspar Schott's *Mechanica Hydraulicopneumatica*, 1657.

Wissenschaften zu Göttingen in 1752. As the University there was founded by our George II in 1734—his bust dominates the Aula—we may take a particular interest in this publication. I mention the appearance of these four journals to remind you that, although the first half of the eighteenth century was not remarkable for great scientific discoveries, yet it saw a great activity in the production of media for their publication.

Through the eighteenth century there continued the lecture, the book and the scientific journal as means of communication. Lectures for interested subscribers became a feature of the cultured life of London: for instance, in 1717, Desaguliers published what he calls in the preface 'Minutes of my lectures, for the Use of such Gentlemen as have been my Auditors'; another book of about this date describes a course of experiments to be performed by Francis Hauksbee, with explanatory lectures read by William Whiston.

The end of the century saw a fresh great development of the channels of scientific information in the periodical limited to a particular branch, or to certain closely allied branches, of science. In 1771 came forth the *Transactions of the American Philosophical Society*, with which event Benjamin Franklin was prominently associated. This publication has continued without a break to the present day. The *Annales de Chimie et de Physique* first appeared in 1789 and was followed by the *Annalen der Physik und der Chemie* (which developed out of the *Journal der Physik*, started in 1790), both of which exist to this day, except that the chemical part of the title has disappeared. Our own *Philosophical Magazine* first came out in 1798. Thus by the end of the century England, France and Germany all possessed journals for the publication of original work in physics, which have continued to the present day.

Another event of importance was the establishment of the provincial scientific society. I do not pretend to have entered upon a careful investigation as to which was the first, for it would be a long and argumentative undertaking. In any case, typical of what I mean is the Manchester Literary and Philosophical Society, whose *Memoirs* came forth in 1789. John Dalton joined the Society in 1794, and became President in 1817, an office which he continued to hold until his death in 1844. Many of his papers, on a variety of subjects, appeared in the *Memoirs* of the Society. In the early years of the present century Rutherford and his collaborators contributed papers and the Society still maintains its high reputation. The Manchester Literary and Philosophical Society is an outstanding example, but in general the local scientific societies fulfilled an important function in the spreading of scientific information, especially in the field of natural history. In those times the country parson was often a well-trained naturalist, sometimes an astronomer, and he did much to foster, through the local society, a taste for science and, occasionally active scientific work. To-day, in general, he has neither the training, the background nor the leisure necessary for such work, and, whether due to this or to other causes, the local societies and clubs seem, unfortunately, to be diminishing in number and activity, although some preserve their status.

The early years of the nineteenth century saw the first appearance of scientific journals in the modern sense in certain countries justly celebrated for the brilliance of their scientific achievements. In 1812 were born the *Proceedings of the Royal Academy of Sciences of Amsterdam*; in 1819 the *Memoirs of the Royal Lombard Institute of Science and Letters* announced the growing activity in Italy. Belgium followed in 1835 with the *Bulletin de l'Académie Royale des Sciences et Belles Lettres de Bruxelles*. In 1842 came forth the *Acta* of the Finnish Society of Sciences. America already had the *Transactions of the American Philosophical Society*, to which reference has been made, but an important event in her scientific history was the initiation of *The American Journal of Science* (1819), often called *Silliman's Journal* since it was founded by the well-known Benjamin Silliman and carried on by his son, and later came the *Proceedings of the American Philosophical Society* (1840), closely followed by the *Proceedings of the American Academy of Arts and Sciences* (1848).

With the nineteenth century the specialized journal, devoted to one branch, often a very limited branch, of science, began its great development, which, I suppose, is the main cause of this conference. To take a few examples, *Crelles Journal*, where so many great mathematical papers are to be found, first appeared in 1826 and *Dinglers Polytechnisches Journal* in 1820. The flood of specialized chemical journals was heralded by Liebig's *Annalen der Pharmazie* (1832), which in 1840 became the *Annalen der Chemie und Pharmazie*. The chemical encyclopaedias, which culminated in the great Beilstein, which surely must come in for discussion at the conference, began, as far as I know, with the *Handwörterbuch der Reinen und Angewandten Chemie*, which likewise owes its inception to Liebig. It would have been extremely laborious for me to amass, and tedious for you to hear, detailed statistics about the journals of the nineteenth century, but, taking as an example those dealing with physics in all its branches—those from which the data in the physical tables of Landolt-Börnstein are compiled—in the period 1810 to 1910 about eighty *new* journals appeared, roughly at the rate of eight new journals every ten years. This is only one branch of science, and only up to 1910. You will, no doubt, hear just what the position is nowadays from those more familiar with the situation than I am.

The foundation of the *Versammlungen Deutscher Naturforscher und Ärzte* in 1822 represents a new channel of scientific communication, both by personal contacts and publication—especially the former—which quickly found imitators. It organized annual meetings of men of science in different German towns for the purpose of discussing the latest advances and also of arousing interest in science in the locality. Our *British Association* followed in 1831, and other countries have similar institutions, such as the *Association Française pour l'Avancement de la Science*.

With the enormous growth of science in the nineteenth century a new portent appeared, the abstracting journal. The Royal Society was early in the field, for the *Proceedings*, which appeared in 1832, were in the first case merely a collection of abstracts. It was not until the seventh volume, of 1856, that short original papers began to appear. The abstracts were not of papers in general, or even of all English papers, but only of those that had appeared in the *Transactions* of the Society. Chemical abstracts first appeared in Germany as the *Pharmaceutisches Central-Blatt* in 1830 and in England in the publications of the Chemical Society in 1841, although they were not published separately until 1878. The *Beiblätter* to the *Annalen der*

Physik began in 1877 the publication of abstracts of all physical papers except those printed in the *Annalen* itself. About the same time appeared the botanical abstracts *Botanischer Jahresbericht* (1874) and the zoological *Zoologischer Jahresbericht* (1880). By the end of the century many other series of abstracts had appeared, including our *Science Abstracts*, and you know, or will soon know, what the situation is like to-day.

Brief mention must also be made of special bibliographies, such as *The Royal Society Catalogue of Scientific Papers*, which covers the whole of the nineteenth century, and, more particularly, of the periodical bibliographies that, especially in the biological sciences, are widely used to-day. Typical are the *Zoological Record*, which began in 1864, the *Index Medicus* (1879) and the *Index Kewensis* (1893).

I cannot leave the nineteenth century without referring to a development in which England led the world, namely, the appearance of the periodical *Nature* in 1869. The objects were announced to be 'First, to place before the general public the grand results of scientific work and scientific discovery; and to urge the claims of science to a more general recognition in education and in daily life; and, secondly, to aid scientific men themselves, by giving early information of all advances made in any branch of natural knowledge throughout the world, and by affording them an opportunity of discussing the various scientific questions which may arise from time to time'. It would be superfluous here to describe the great part which *Nature* has played and now plays in the communication of scientific information. It is, for many purposes, an international organ which is unique. Where the vexed question of preserving priority arises, a letter to *Nature* is everywhere the solution. In the seventeenth century it was the custom to publish, in the interest of priority, an anagram in which the letters of a significant sentence were arranged in alphabetical order. The best-known example is probably Hooke's *ceiiinosssttuv* which, properly arranged, makes *ut tensio sic vis*, Hooke's law. The anagram was, of course, unintelligible until the author was challenged by event and construed its meaning. Some writers of letters to *Nature* have a strong historical sense.

It would not be right to omit the book as a channel of original scientific information and a valuable aid and incitement to original research. I have already referred to important early books, such as Copernicus's *De Revolutionibus* and Gilbert's *De Magnete*. These books were widely read—Copernicus, for instance, was printed four times within the hundred years that started with its first appearance, and Gilbert four times within forty years. The influence of Newton's *Principia*, of which three editions appeared in forty years, and of his *Opticks*, of which four editions appeared within thirty years, it would be hard to overrate. I have already noted that in the seventeen years which elapsed between the books Newton had abandoned Latin for first publication and passed to the vernacular, although the *Opticks* was speedily translated into Latin for the benefit of the Continent. He was not, of course, by any means the first to publish in English: Robert Boyle's books all appeared first in English and then in Latin, and there are important Elizabethan scientific books in English. Huygens published his early books, such as the *Horologium Oscillatorium* (1673), in Latin, but the *Traité de la Lumière* appeared in 1690 in French. Thus the turn of the seventeenth century heralded the passing of Latin as a universal language, although for the next fifty years it is common to find books in that tongue, for instance, those of the Bernoullis and the Eulers. With the general decline of the classical influence dwindled also the use of the dialogue form, no doubt a memory of Plato. Examples of the use of this form are Fernel's *De Abditis Rerum Causis* and Galileo's *Dialogo*, in which there debate Salviati, a Copernican; Sagredo, a witty and educated scholar, desirous to learn; and Simplicio, an Aristotelian who puts forward the stock arguments of the scholiasts in an absurd way. You know, of course, the tradition that Galileo put certain arguments which the Pope had suggested to him into Simplicio's mouth, and that that was the real cause of the trouble into which Galileo fell. The last great book that I can recall in dialogue form is Boyle's *Sceptical Chymist*.

Books that have made scientific history will readily occur to everyone. As examples we may take Linnaeus's *Genera Plantarum*, Lagrange's *Mécanique Analytique*, Lavoisier's *Traité Elémentaire de Chimie*, Carnot's *Puissance Motrice du Feu*, Liebig's *Organische Chemie*, and in more modern times Clerk Maxwell's *Electricity and Magnetism*, J. J. Thomson's *Conduction of Electricity through Gases*, Rutherford's *Radioactivity*, Sherrington's *Integrative Action of the Nervous System* and Rayleigh's *Sound*. You will have to consider, I am afraid, the book and no doubt you will be urged from some quarters that all books ought to be the same size. It would probably save wood in bookshelves, or steel, and is, I suppose, by no means a matter to be hastily dismissed by questing minds. Some forward-thinking spirits may further suggest that they should all be printed by the same one publishing house, which by due coordination, a word to which Professor Rocard has recently devoted a learned treatise, would—should, rather—lead to great economies, as production would then be properly planned.

There are certain questions of the modern production of scientific books which are raised by the study of the old books of science. The beauty of the old production we can scarcely hope to imitate nowadays: craftsmen like those who, cutting the woodblocks for geometrical diagrams, found the work dull and so filled in the blank spaces with elegant little designs, producing the so-called floriated diagrams, no longer exist. Our materials will not allow us to use the device adopted in Newton's *Opticks* and many books of Newton's time* to make diagrams visible to the reader without turning back while he reads the several pages that contain references to them, namely, to print the diagrams on a folding plate with a blank space, the size of a page, where it is attached to the book, so that when the plate is extended and the book closed the whole engraved part of the sheet is still visible. For one thing, our paper would not stand the repeated folding and unfolding. Descartes solved the same problem of enabling the reader to have always before him a diagram to which reference is made on several pages, by reprinting it on every relevant page—thus in his *Principia* one particular diagram is printed eleven times. We have not the paper for this extravagance.

The old books raise other questions of illustration. They used, in the best period, the woodcut and the copperplate: illustrations such as those to Senguerd's

---

* The earliest example that occurs to me is Lana's *Prodromo*, of 1670, but quite likely this is not the first.

Philosophia Naturalis are little works of art, worthy to be compared with Della Bella's fine frontispiece to Galileo's *Dialogo*. Many aspects of the graver's art appear: thus the illustrations to Thomas Wright's *New Hypothesis of the Universe* (1750) are in mezzotint, which beautifully renders the soft blackness of the night against which the stars shine. To-day we have neither time, money, paper nor craftsmen for illustration on the lines shown in early books, but we have something which our ancestors lacked, photography and the various photographic processes of reproduction. Photography, it must be remembered, is a convention just as much as line drawing: what is produced is a version of the object or phenomenon photographed, and in no sense a true, in the sense of a unique, representation. To realize this, one has only to think of the effect of the direction of illumination on what is seen in a microphotograph of a metal surface, say, or of the differences that can be produced by colour filters and different types of photographic plate in rendering any subject. The use of the photograph has, of course, transformed observational astronomy and bacteriology, and no book on these subjects can be without photographic illustrations. It would be out of place to attempt to trace the manifold uses of the photographic illustration, but it may be permissible to point out that there are many purposes for which a good drawing is preferable. Photographs of instruments and complicated installations of apparatus are seldom successful: it is nearly always inessentials that are prominent in them and awkward shadows have a way of hiding what is wanted. I venture to give here a reproduction of a good photograph of an experimental set-up (figure 4, plate 1): essential features, as, for instance, the form of the arm just above the vessel at the bottom of the picture, hardly tell at all. This should be contrasted with any good line illustration of apparatus, either in old or newer books.* Perhaps I may quote here what Fuchs says in 1542 in the preface to his *De historia stirpium*: 'Furthermore we have purposely and deliberately avoided the obliteration of the natural form of the plants by shadows, and other less necessary things, by which the delineators

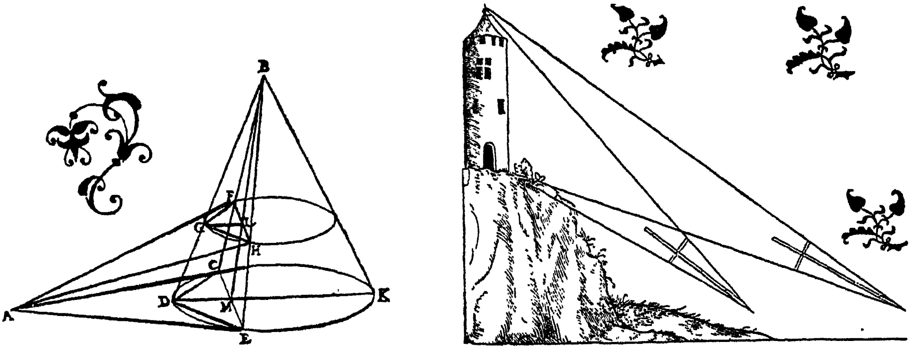


FIGURE 2. Floriated diagrams.

* In my lecture I took as an example of the confused masses frequently shown in photographs of apparatus a picture from a published book; this was much worse than the one here reproduced, which is from a photograph taken in my own laboratory. I can, however, hardly ask the publisher to allow me to reproduce it as a warning.

sometimes try to win artistic glory: and we have not allowed the craftsmen so to indulge their whims as to cause the drawing not to correspond accurately to the truth.' For the representation of plants in their surroundings, ecology, photography is often excellent: for the representation of structural botany the illustrations of a hundred years or more ago have never been equalled. Photographic reproductions need special, so-called art, paper, which is unsuitable for ordinary book pages—it is too heavy and does not make a good page. Incidentally, the weight of books is a matter worthy of some discussion—a very distinguished man of science has protested to me about the great weight of some single volumes printed on art paper and quoted a recent American publication which weighs about 4½ lb. Certain German books in which sections of art paper are introduced among normal paper and printed as ordinary pages have an unpleasant appearance. The whole question of the illustration of scientific books and periodicals is worthy of consideration, and clearly cannot be without influence on format and make-up.

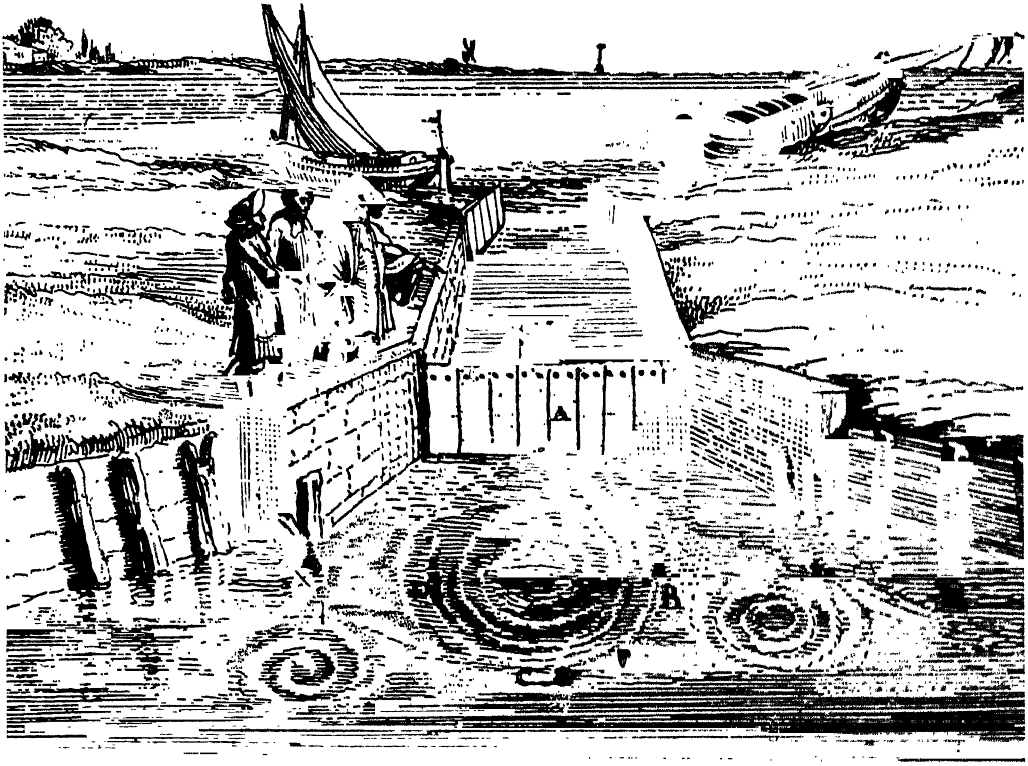


Figure 3. An illustration, by Schoonebeck, from Senguerd's *Philosophia Naturalis*, 1681.

If the old books and journals can offer us an example in many matters of material production, which we may be kept from following by material dearth, they are no less a pattern in certain matters of literary presentation. I would not suggest, of course, that the form of Newton's *Principia* is one that we should imitate to-day, although Thomas Young set out his Bakerian Lecture, in which he developed the wave theory of light, in *Propositions* and *Corollaries*, somewhat in the Newtonian method. It would be vain to wish to hear again the sonorous swing of Georgian prose in science, although rebukes such as 'We can only regret that a person so void of a sense of physical elegance should have an opportunity of obtruding opinions like these on the public' or 'The reviewer has afforded me, in the next observation, an opportunity for a triumph as gratifying as any triumph can be where the enemy is so contemptible' would add to the interest of scientific controversy. I could, however, wish that the clarity and elegance with which so many of the older writers set out their thoughts and experiments would find more imitators to-day. The great papers of J. J. Thomson and of Rutherford have a straightforward lucidity of phrase. When, as a young man, I had to learn something of the calculus of variations, someone suggested that I should read Lagrange's original memoir, and I still remember the pleasure that his exposition gave me. Thomas Young, whom I have just quoted, sets a good example, but he was a great classical scholar, bred in the traditions of literary style. Faraday, however, was practically untaught—'My education was of the most ordinary description, consisting of little more than the rudiments of reading, writing and arithmetic at a common day-school'—and yet his writing of matters sufficiently abstruse shows a beautiful clarity and simplicity. Consider the following passage from his paper in the *Philosophical Magazine* entitled 'A Speculation touching Electric Conduction and the Nature of Matter':

> But it is always safe and philosophic to distinguish, as much as is in our power, fact from theory; the experience of past ages is sufficient to show us the wisdom of such a course; and considering the constant tendency of the mind to rest on an assumption, and, when it answers every present purpose, to forget that it *is* an assumption, we ought to remember that it, in such cases, becomes a prejudice, and inevitably interferes, more or less, with a clear-sighted judgment. I cannot doubt but that he who, as a wise philosopher, has most power of penetrating the secrets of nature and guessing by hypothesis at her mode of working, will also be most careful, for his own safe progress and that of others, to distinguish that knowledge which consists of assumption, by which I mean theory and hypothesis, from that which is the knowledge of facts and laws; never raising the former to the dignity or authority of the latter, nor confusing the latter more than is inevitable with the former.
>
> Light and electricity are the two great searching investigators of the molecular structure of bodies....

This deals with a matter which nowadays is often smothered in a complexity of long words and cumbrous and ambiguous phrases, yet here the sense is clearly and unmistakably expressed, whether some of our moderns would agree with it or not.

I cannot help feeling that much of the obscurity that hangs about many papers to-day is unnecessary, and I am in good company in objecting to it, for Lord Rayleigh records of J. J. Thomson 'Like Lord Kelvin he was in general impatient of obscurity'. I am not thinking of our Newtons: they must go their own way, although we may note that William Molyneux writing to Hans Sloane in 1697 said: 'I hear Mr. Newton's *Phil. Nat. Prin. Math.* is out of press, and that he designs a 2nd Edition. Pray advise him to make it a little more plain to Readers not so well versed in Abstruse Mathematicks, a few Marginal Notes and references and Quotations will doe the business.' The gentlemen whom I find tiresome are those who write matter that is unintelligible not on account of its difficulty but on account of the confusion of thought, the lack of sequence and the tangled, sometimes ungrammatical jargon in which it is expressed. Again, there are certain writers of papers on mathematical physics whose productions are difficult to appreciate because the assumptions which have been made are never clearly expressed, but introduced casually by the application of formulae designed for other circumstances. Others have an annoying habit of complicating matters by introducing a spurious and superficial generality which is thrown away as soon as the real problem which is under treatment is approached. It would save the reader much trouble if every writer of a scientific paper would first state clearly what he has done and why it is of interest: if a mathematician would say exactly what he has assumed, what method he has applied and what results obtained. I have sometimes wondered if certain important mathematical papers dealing with physical problems would not have a much wider circle of readers—and of readers who would profit—if much of the work were relegated to a kind of mathematical appendix, the writer saying that from one stage he has passed to the next by the process described in this appendix. I realize that I am laying myself open to envenomed attack, but the attackers may even learn something in sitting down to explain just how absurd my suggestion is.

Many who realize how much some of the papers written to-day, especially as first presented, are wanting in clarity and conciseness, to say nothing of elegance, have suggested that our young research workers should be made to take courses in English in the English departments of our universities. I scarcely think that this is the way to go about it. Some of the worst English that I have read has been written by D.Litt.s, and I used to cut out the writings of a certain headmaster that occasionally appeared in the newspapers to keep them as examples of turgid language. Rather read to the young man an account of a piece of work as written by a master and then let him write his own account and compare it subsequently with the original. Personally I feel that a wrong ideal is often the cause of much of the trouble. The young man wants a long paper suspecting, in many cases rightly, that when it comes to applying for a higher post his achievement will be estimated by the number of pages that he has produced. He wants one that looks as formidable as possible judging, in many cases rightly, that the senior who looks at it will think more highly of him if it is full of forbidding phrases and long words, especially if they are new words, such as, let us say, diclastically disintegrated to mean 'cut into two bits'. If it were conveyed to him that simple expression is in itself desirable, that every new word needs careful justification, that scientific jargon is not in itself scientific advance there would, I think, be an improvement. Personally, when a young man brings me a mass of verbiage I ask him, gently, what exactly he means, and when, as usually happens, after much questioning he has reached a perfectly simple form of words I say 'Then why not say so?' I think that it is as simple as that. But, to use the language of to-day, I am sticking out my neck. Perhaps I may save myself some blows in this vulnerable region if I insist that I realize that everybody cannot write great and memorable prose like that of Charles Sherrington —but then he is a fine poet and for fine poets words have a ring, a friendly aptness like that which his tools have for the master craftsman, which they do not have for the actuary. All that I ask is a freedom from unnecessary clumsiness and obscurity. I do not ask that scientific papers shall be pleasant to read, only that they shall not be repugnant.

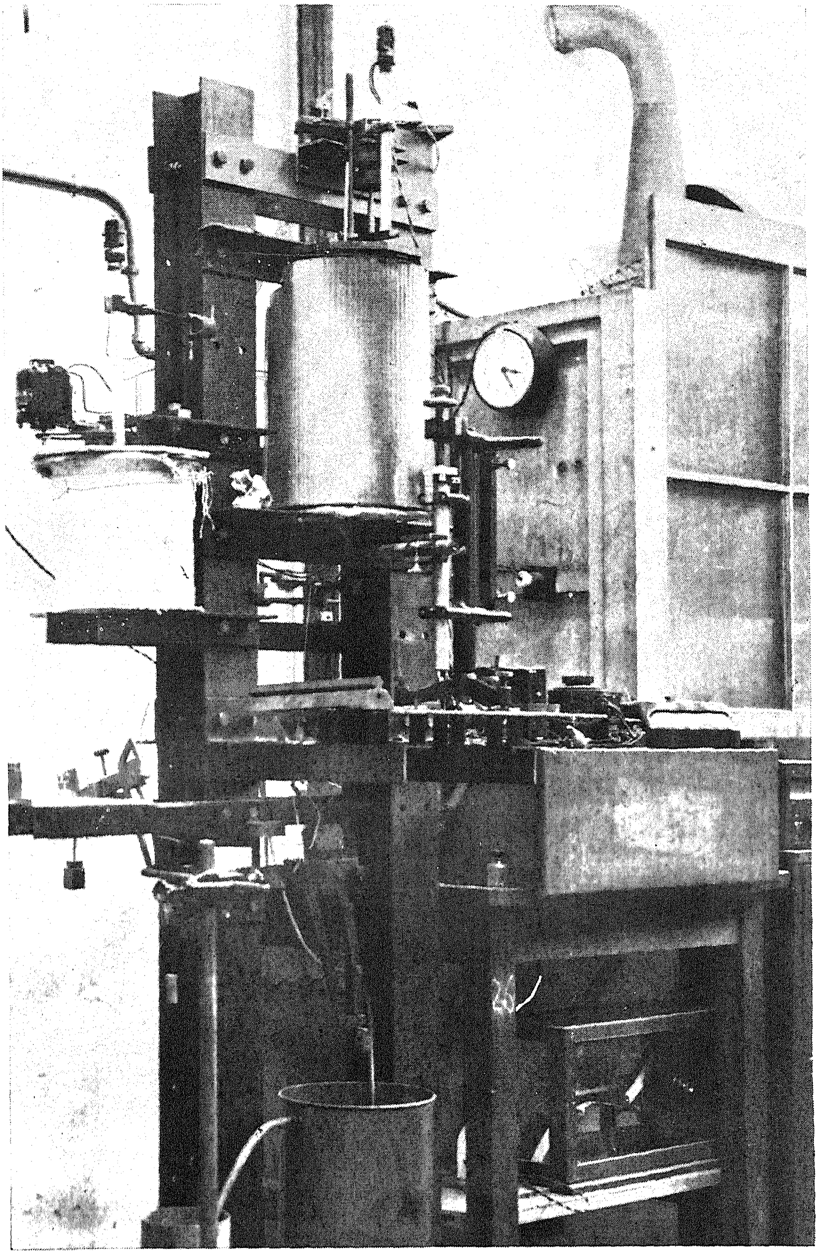


FIGURE 4. A photograph of a simple assembly of apparatus.

There is to-day a growing specialization in science which is inevitable. We no longer have men of learning writing for other men of learning, or men of science writing for other men of science, or disciples of exact science writing for other disciples of exact science, or physicists writing for other physicists, or students of optics writing for other students of optics, or spectroscopists writing for other spectroscopists, but rather infra-red spectroscopists writing for other infra-red spectroscopists—and very soon they will have a journal of their own. You, gentlemen, are dealing with the very complex problems that such a situation creates, and I am confident that great simplification and great economies will result from your labours, and that the toil of the specialists in the smaller and smaller regions of greater and greater complexity will be lightened. But the great advances will still be made by solitary spirits who have greatly pondered on more general problems, who have, like Newton, kept the subject constantly before them until the first dawnings open little by little into the full light. Let us take care that we do not organize things to such a point that no leisure exists, since, for the pioneers, leisure is even more needed than journals, reference books and resolutions.

For certain periodicals which are in an unfamiliar language, e.g. Swedish, I have given the English equivalent of the title. Readers who want the actual title can find it in the list of periodicals given at the end of any edition of the Landolt-Bornstein *tables.*

On the bond lengths and interatomic distances in certain molecules and crystals

By W. Hume-Rothery, F.R.S.
Inorganic Chemistry Laboratory, University Museum, Oxford

(*Received* 3 *November* 1948)

The internuclear distances in the crystals of some of the elements at the absolute zero have been calculated, and the values are examined and compared with those in some molecules. In the first two short periods if the single bond length is denoted D, and the atomic number Z, a direct proportionality exists between D^2 and Z for the members of any one group from group IV to group VII. In groups IV to VII this proportionality extends to the corresponding elements of the first long period, so that, for example, the D^2 values for carbon, silicon and titanium are proportional to Z. In group III the D^2 value for boron and the square of the internuclear distance in the crystal of aluminium are proportional to Z, but the proportionality does not extend to scandium, and in groups I and II no proportionality exists between D^2 and Z. Simple whole-number relations are found between the squares of the internuclear distances in the earlier members of group II, and other regularities are pointed out. Evidence is also given for the hypothesis that simple whole number relations exist between the squares of the bond lengths of single, double and triple bonds of a given element.

1. Introduction

The modern electron theory has reached the stage at which the interatomic distances in some relatively simple crystals and molecules can be calculated approximately, but in many cases the problem is too difficult for accurate solution. It is therefore of interest to see whether the experimental data suggest any underlying regularities, since although such relations may be empirical they may assist the development of the theory. In a previous paper (Hume-Rothery 1930) simple relations were pointed out which held approximately between the interatomic distances in the crystals of some of the elements and the atomic numbers of the elements concerned. In this paper the interatomic distances were estimated at one-half the characteristic temperature, and the relations were only approximate, the deviations being of the order of several per cent. It seemed probable that any really fundamental relation would refer to the interatomic distances at the absolute zero, and in a later paper (Hume-Rothery 1945) the use of Grüneisen's equation for calculating changes in volume has been discussed. The results obtained have shown that reduction to the absolute zero does not remove the discrepancies between the data and the relations referred to above, but in the course of this work, new and more accurate relations have been discovered and these are described in the present paper.

2. GENERAL

On proceeding down the Periodic Table of the elements the continuous increase in the atomic number Z results in a general contraction in the electron cloud of any one group or subgroup of electrons. On passing from hydrogen to calcium we expect therefore a general contraction broken by abrupt expansions at lithium, sodium and potassium where new quantum groups are opened up. We may therefore hope to find relations between the interatomic distances in the elements of any one group in the first and second short periods and the first long period because each step (e.g. $C \rightarrow Si \rightarrow Ti$) involves an increase of 8 in the atomic number. We shall not, however, expect the same regular variation on passing to the second long period because this involves an increase of 18 in the atomic number.

3. ACCURACY OF EXPERIMENTAL DATA

The accuracy with which the interatomic distances at the absolute zero can be estimated varies greatly. In the cubic metallic structures where the interatomic distances can be calculated directly from the lattice spacings, the room temperature measurements are usually highly accurate (1 part in 10,000 to 1 part in 50,000). For the face-centred cubic and body-centred cubic metals the values at the absolute zero should be reliable to 1 part in 500, except for the alkali and alkaline earth metals of higher atomic number. For the close-packed hexagonal metals the accuracy is less because the axial ratio has to be determined,* and the reduction to the absolute zero involves the ratio of the coefficients of expansion parallel and perpendicular to the hexagonal axis and the ratio varies with the temperature; the variation is known for magnesium, zinc and cadmium but not for most other metals. In the more complicated crystal structures, where the interatomic distances are deduced from the intensities of X-ray reflexions, the accuracy is much less, and errors of the order 1 to 3 % may be expected. In such cases one cannot in general calculate* the contraction which would take place on cooling to the absolute zero, but where simple covalent bonds are concerned† the theory indicates that it will be relatively small. Where interatomic distances in molecules are determined by electron diffraction methods the results are subject to errors of a few per cent, and it is again not possible to say how much these‡ would be affected by reduction to the absolute zero, although the change is not likely to exceed a few per cent. Interatomic distances in molecules obtained by the analysis of molecular spectra are generally expressed in terms of r_e, the equilibrium distance between two atoms in the absence of vibration. This distance corresponds to the minimum on the curve connecting the potential energy and the internuclear separation, and it is the r_e values which are usually calculated theoretically. In any actual molecule there is always some vibration, and the mean distance between two atoms in the lowest vibrational state is denoted r_0. The value r_0/r_e approaches unity as the atomic weights of the atoms increase. Thus in the H_2 molecule $r_0/r_e = 1{\cdot}013$, whilst in the Cl_2 molecule $r_0/r_e = 1{\cdot}002$. In comparing the interatomic distances in metallic crystals with those in molecules, the r_0 values should really be used, but the difference is usually too small to be significant for the present paper. Provided that the spectroscopic data are analyzed correctly§ the accuracy of spectroscopic r_e values is much greater than those obtained by electron diffraction methods, and may be taken to be of the order 1 part in 500 with greater accuracy in some cases. It must be remembered that the true Ångström unit values of the spectroscopist require dividing by 1·00202 to be converted to values in the kX units of the X-ray crystallographer.

---

* In first-class lattice spacing work the lattice spacings are of course determined accurately, but for many elements it is clear that the data for close-packed hexagonal structures are less accurate than for the body-centred or face-centred cubic structures.

4. HYDROGEN, HELIUM AND THE ELEMENTS OF GROUPS III TO VI

We may consider first the elements in the later groups of the first two short periods. The symbol D may be used for the length of the single covalent bonds in the F_2 and Cl_2 molecules, the former being known from electron diffraction methods, and the latter from spectroscopic determinations. We then find that D^2 for these two elements is proportional to the atomic number Z. This is shown by the lowest line in figure 1, and the relation holds to within the accuracy with which the data are known. The same figure shows that the straight line from the origin through the points for F and Cl passes very nearly through the D^2 point for the closest distance of approach in β-manganese. This may perhaps support the views of Pauling (1938) that the cohesive forces in the crystals of the transition elements resemble those of covalencies.

---

* In a complicated structure such as that of manganese, the coefficient of thermal expansion enables the change in lattice spacing to be calculated, but the parameters which determine the positions of the atoms may change with temperature.

† The author has to thank Professor C. A. Coulson for helpful discussion on this point.

‡ Some electron diffraction results refer to vapours at the boiling-points of the substances concerned.

§ The correctness of the analysis is of course essential; thus in the case of ethylene, published values for the length of the C=C bond lie between 1·33 and 1·37 Å, and the differences are the result of incorrect assumptions introduced into the calculations.

In group VI the diatomic molecules O_2 and S_2 involve double bonds, but the single-bond lengths have been determined for oxygen by electron diffraction studies of H_2O_2, and by the X-ray determination of the crystal structure of hyperol, whilst for sulphur the single-bond lengths have been determined by electron diffraction studies of S_2Cl_2, liquid sulphur and plastic sulphur, and by the X-ray crystal analysis of rhombic sulphur. As will be seen from figure 1, D^2 is again proportional to Z within the limits of accuracy of the data, and the straight line on extrapolation passes near to the point for the D^2 value of α-chromium.

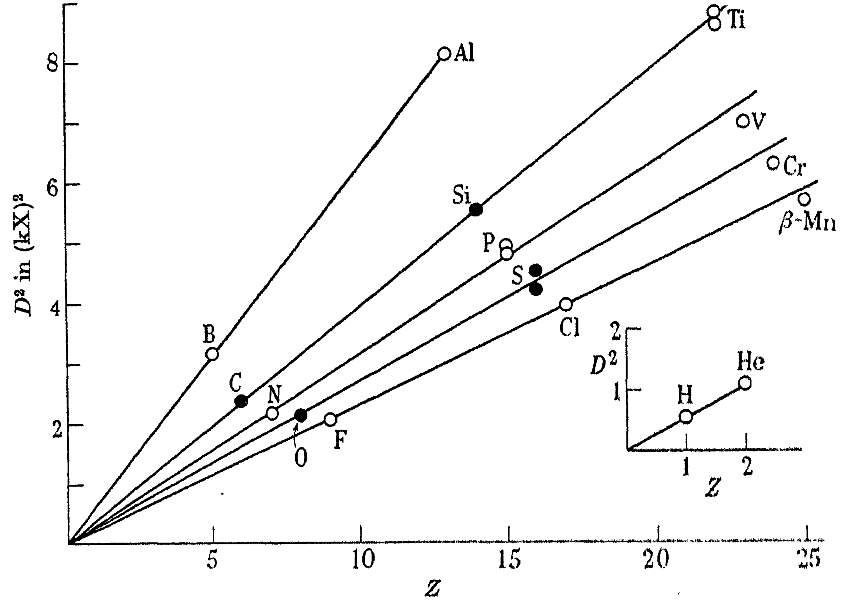


FIGURE 1

In group V the single-bond lengths have been determined for nitrogen by electron diffraction studies of hydrazine, and for phosphorus by electron diffraction studies of the P_4 molecule, and by X-ray crystal analysis of black phosphorus. As shown in figure 1, there is again a direct proportionality between D^2 and Z, and the straight line on extrapolation passes slightly above the point for vanadium. It will be seen that in groups V and VI the points for vanadium and chromium lie slightly below the straight lines through the points for the earlier members of the groups, and this may well be the result of the fact that the points for these metals have been corrected so as to refer to the absolute zero, whilst a similar correction cannot be made for the electron diffraction results or for the complicated crystal structures of sulphur and phosphorus.

In group IV the D^2 values in crystals of carbon, silicon and titanium are very nearly proportional to Z as can be seen from figure 1. In the case of the diamond, the reduction to the absolute zero can be made with accuracy, but the result given for silicon refers to −150° C, because the thermal expansion at lower temperatures is abnormal, and a negative coefficient of expansion has been claimed. It is improbable that this would raise the value for silicon at 0° K sufficiently to give an exact proportionality between D^2 and Z, but, as can be seen from table 1, even if the value for silicon at $-150°$ C is compared with that for the diamond at 0° K the departure from an exact proportionality is only of the order 3 parts to 400 on D^2 or *c.* 0·4 % on D. Titanium crystallizes in the close-packed hexagonal structure with axial ratio $c/a = 1{\cdot}601$, and each atom has six neighbours at a distance a and six at a distance equal to $\sqrt{\left(\frac{a^2}{3}+\frac{c^2}{4}\right)}$. The corresponding D^2 values have been inserted in figure 1, and the straight line through the points for carbon and silicon passes between them.

TABLE 1

element and atomic number		D (kX units)	D^2	D^2/Z	remarks
F	9	1·43(5)	2·05(9)	0·229	electron diffraction of F_2
Cl	17	1·980	3·920	0·230(6)	spectroscopic Cl_2 molecule, r_e value
β-Mn	25	2·365	5·593	0·224	X-ray crystal analysis of complex structure at room temperature
O	8	1·47	2·16	0·270	electron diffraction of H_2O_2 at 150° C
		1·46	2·13	0·266	X-ray crystal structure analysis of hyperol
S	16	2·12	4·49	0·281	X-ray crystal structure of sulphur at room temperature
		2·08	4·33	0·271	electron diffraction of plastic sulphur
		2·07	4·28	0·268	electron diffraction of liquid sulphur
		2·04	4·16	0·260	electron diffraction of S_2Cl_2
α-Cr	24	2·490	6·200	0·258	X-ray crystal structure at 0° K
N	7	1·47	2·16	0·309	electron diffraction of hydrazine vapour
P	15	2·18	4·75	0·317	X-ray crystal structure of black phosphorus at room temperature
		2·21	4·88	0·325	electron diffraction of P_4 molecule
V	23	2·623(4)	6·882	0·299	X-ray crystal structure at 0° K
C	6	1·5412	2·3753	0·3959	X-ray crystal structure of diamond at 0° K
Si	14	2·345(1)	5·499(5)	0·3928	X-ray crystal structure at $-150°$ C
Ti	22	2·949 2·911	8·697 8·474	0·395 0·385	two closest distances of approach in close-packed hexagonal structure at 0° K
		2·930	8·584	0·390	D_i value, see p. 22
B	5	1·78	3·17	0·634	single-bond length, see p. 22
Al	13	2·8454	8·0963	0·623	X-ray crystal structure at 0° K
H	1	0·7399	0·5475	0·5475	H_2 molecule, spectroscopic, r_e value
He	2	1·044	1·090	0·545	He_2 molecule, spectroscopic, r_e value

The above data suggest clearly that in the elements of groups IV to VII in the first two short periods, it is a general principle that D^2 is directly proportional to Z for the elements of any one group, and that the same proportionality holds approximately when extended to the corresponding elements of the first long period.

In group III the single-bond length of boron has been in dispute because the B-B distance in the B_2H_6 molecule which was originally thought to be that of a normal single bond is now generally regarded (Pitzner 1945) as involving a 'protonated' double bond. According to Levy & Brockway (1937) the normal single-bond length of boron in a compound where the boron atom is surrounded by an octet of electrons, that is to say, a compound in which all four orbitals are used, is 1·78 kX, and this gives a value of $D^2 = 3{\cdot}17\ (\text{kX})^2$. This value has been included in figure 1, from which it will be seen that the points for boron and aluminium indicate a direct proportionality between D^2 and Z. According to Levy & Brockway, in a compound where the boron atom is surrounded by six electrons and only three orbitals are used, the B-B single-bond length is only 1·58 kX, and the corresponding value of $2{\cdot}50\ (\text{kX})^2$ for D^2 is too low to give a direct proportionality in figure 1. This may again support the views of Pauling, who regards the cohesive forces in aluminium as resulting from resonance involving four orbitals per atom.

Figure 1 shows also that a direct proportionality exists between D^2 and Z for the interatomic distances in molecules of H_2 and He_2, and as will be seen from table 1 the relation is very precise. This is of interest because the He_2 molecule concerned is not formed by the van der Waals forces between neutral atoms, but is a molecule whose ground state is derived from helium atoms in an excited state, presumably involving electrons in two-quantum orbitals. This means that in passing from H_2 to He_2 the principal quantum number of the bonding electrons increases by one, and the condition of affairs is thus in some respects analogous to that on passing from F_2 to Cl_2.

5. THE ELEMENTS OF GROUP II

Of the earlier elements in group II, calcium and strontium crystallize in the face-centred cubic structure, and the closest distances of approach (D) at the absolute zero can be calculated, although the accuracy is not very high because the coefficients of expansion have not been fully investigated.* Beryllium ($c/a = 1{\cdot}5682$) and magnesium ($c/a = 1{\cdot}62354$) crystallize in the close-packed hexagonal structure with axial ratios not very different from that ($c/a = \sqrt{\frac{8}{3}} = 1{\cdot}633$) for close-packed spheres. This results in each atom having twelve close neighbours, of which six are at one distance and six at a slightly different distance. In this case we have found it convenient to calculate an ideal distance of approach D_i which, if the axial ratio were equal to the ideal value of $\sqrt{\frac{8}{3}}$, would give the same volume per atom Ω as that found in the actual structure. This policy may be criticized, but it seems a reasonable way of comparing the interatomic distances in face-centred cubic and close-packed hexagonal structures. The same policy has been adopted for zinc ($c/a = 1{\cdot}8563$ (1)), where the close-packed hexagonal structure is elongated in the direction of the c-axis. The reduction to the absolute zero can be made accurately for all three metals, although the accuracy is slightly less for zinc owing to the curious effects at very low temperatures where the coefficient of expansion perpendicular to the hexagonal axes changes sign. It will be appreciated that a comparison of D for the face-centred cubic structure with D_i for the close-packed hexagonal structures is equivalent to a comparison of $\Omega^{\frac{1}{3}}$, and a comparison of D^2 and D_i^2 to one of $\Omega^{\frac{2}{3}}$. For the body-centred cubic structure the relation between D and Ω is of course different.

In figure 2 the values of D_i^2 for beryllium and magnesium, and of D^2 for calcium and strontium are plotted against the atomic number Z. It will be seen that in contrast to the behaviour in groups III to VII there is no proportionality between D_i^2

---

* In the absence of data for strontium the coefficient of expansion was assumed to be the same as that for calcium.

and Z. The points for beryllium, magnesium and calcium do, however, lie accurately on a straight line,* and further they are in the ratio of the small whole numbers 1 : 2 : 3 to an accuracy which is of the same order as that with which the data for the absolute zero are known (see table 2). On proceeding further we find that

$$D_i^2(\mathrm{Mg}) : D_i^2(\mathrm{Zn}) = \tfrac{1}{3} : \tfrac{1}{4}, \quad D^2_{\mathrm{Ca}} : D^2_{\mathrm{Sr}} = \tfrac{1}{6} : \tfrac{1}{5}.$$

These whole-number relations hold to the same degree of accuracy as that of the data for the absolute zero, and they suggest that the D^2 and D_i^2, and hence the $\Omega^{\frac{2}{3}}$, values for the elements concerned are all related by small whole numbers.

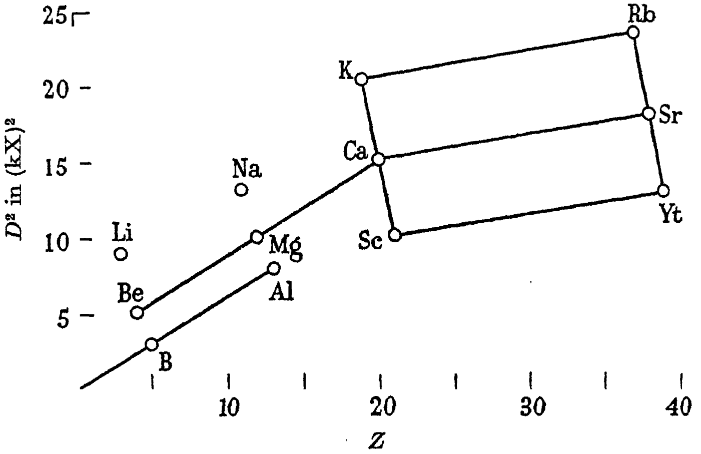


FIGURE 2

TABLE 2

element and crystal structure	D or D_i (kX units)	D^2 or D_i^2	remarks
Be c.p.hex.	2·2476	5·052	—
Mg c.p.hex.	3·1806	10·116	10·116 ÷ 2 = 5·058 as compared with 5·052 for Be
Ca f.c.c.	3·911	15·296	15·296 ÷ 3 = 5·099 as compared with 5·052 for Be
Zn c.p.hex.	2·7566	7·599	7·599 × $\frac{4}{3}$ = 10·132 as compared with 10·116 for Mg
Sr f.c.c.	4·27(3)	18·2(6)	18·26 × $\frac{5}{6}$ = 15·22 as compared with 15·29 for Ca

All values in table 2 refer to 0° K.

6. THE ELEMENTS OF GROUPS I, II AND III

Lithium, sodium, potassium and rubidium crystallize in the body-centred cubic structures, and the corresponding values of D^2 have been plotted in figure 2, which shows also the D^2 value for aluminium and scandium (face-centred cubic), and the D_i^2 value for yttrium (close-packed hexagonal). The point for calcium lies midway between those for potassium and scandium, and further the values of D_i^2 for these three elements are in the ratio $\frac{1}{3} : \frac{1}{4} : \frac{1}{6}$ to a degree of accuracy which is comparable with that of the data at the absolute zero (see table 3). In the next row of the Periodic Table the point for strontium lies midway between those for rubidium and yttrium to within the limits of accuracy of the data, but the values are no longer in the ratio of small whole numbers. The fall in D^2 on passing from potassium to scandium is equal to that on passing from rubidium to yttrium to within the accuracy of the data. It appears, therefore, that in these two rows of the Periodic Table the D^2 values decrease in equal steps on passing from group I to group II and from group II to group III, and that the steps are the same in each period. Scandium (1200° C) and yttrium* (1500° C) are elements of high melting-points which indicate a firm atomic binding. In contrast to this aluminium (660° C) has a comparatively low melting-point (i.e. a weaker binding), and as will be seen from figure 2 the point for aluminium is higher than would be required to give equal falls in D^2 on passing from $Na \rightarrow Mg \rightarrow Al$. This difference may be connected with the fact that in the series $Na \rightarrow Mg \rightarrow Al \ldots$ the bonding is of the (*sp*) type, whereas in the long periods (*spd*) bonding is formed. The interesting fact then emerges that whereas the D^2 values for calcium and scandium are in the ratio $\frac{1}{4}:\frac{1}{6}$, the D_i^2 and D^2 values for magnesium and aluminium are accurately as $\frac{1}{4}:\frac{1}{5}$ (see table 3). These are two of the elements for which the reduction to the absolute zero can be made with the greatest accuracy, and the agreement with the whole-number relation is exact to within 1 part in 2000.

---

* This straight line is nearly but not exactly parallel to that through the points for boron and aluminium.

TABLE 3

element and crystal structure	D or D_i (kX units)	D^2 or D_i^2	remarks
Li b.c.c.	3·009	9·054	—
Na b.c.c.	3·654	13·35	$3D^2 = 40{\cdot}05$
Mg c.p.hex.	3·1806	10·116	$4D_i^2 = 40{\cdot}46$
Al f.c.c.	2·8454	8·0963	$5D^2 = 40{\cdot}48$
K b.c.c.	4·533	20·55	$3D^2 = 61{\cdot}65$
Ca f.c.c.	3·911	15·30	$4D^2 = 61{\cdot}20$
Sc f.c.c.	3·197	10·219	$6D^2 = 61{\cdot}31$
Rb b.c.c.	4·85(1)	23·5	—
Sr f.c.c.	4·27(3)	18·2(6)	$5D^2_{Sr} = 91{\cdot}3$ $6D^2_{Ca} = 91{\cdot}8$
Yt c.p.hex.	3·62(2)	13·1	—
Cs b.c.c.	5·21	27·1(4)	$6D^2_{Cs} = 162{\cdot}(8)$ $7D^2_{Rb} = 164{\cdot}5$ $8D^2_{K} = 164{\cdot}4$

All values in table 3 refer to 0° K.

If the relations between the D^2 values for sodium and magnesium were analogous to those between potassium and calcium we should expect $D^2_{Na} : D^2_{Mg} = \frac{1}{3} : \frac{1}{4}$, and as will be seen from table 3 this is the case to within 1 %. Unless sodium behaves abnormally at very low temperatures this difference is probably outside the experimental error, because the coefficient of expansion has been measured between room temperature and 80° K. The fact that the D^2 value for sodium is just too small to fit the above relation prevents any simple relation being found between the D^2 values for the alkali metals as a whole, although as can be seen from table 3 $D^2_{\mathrm{K}}:D^2_{\mathrm{Rb}}:D^2_{\mathrm{Cs}} = \frac{1}{8}:\frac{1}{7}:\frac{1}{6}$ to within the accuracy with which the data are known.*

---

* The melting-point of yttrium is given as 1490 ± 200° C in the *Metals Handbook*, 1948, p. 21, and is thus very uncertain.

7. RELATIONS FOR MULTIPLE BONDS

The relations described in the preceding sections suggest that simple relations are found if the values for the squares of the interatomic distances are compared, and it is therefore of interest to see whether analogous relations exist between the D^2 values for the single-, double- and triple-bond lengths of a given element.

Table 4 shows that the D^2 values for the single O—O and double O=O bonds are in the ratio of $\frac{1}{2}:\frac{1}{3}$ to within the limits of accuracy of the experimental data. A similar relation exists between the D^2 values for the N—N and N=N bonds, whilst the D^2 value for the N≡N bond is only very slightly larger than that required to give a simple $\frac{1}{2}:\frac{1}{3}:\frac{1}{4}$ series for single-, double- and triple-bond D^2 values; the difference is within that of the experimental accuracy for the N—N and N=N bond lengths. It is also of interest to note that the D^2 value for the N≡N bond is one-third of the D^2 value for the P≡P bond, whilst the D^2 values for P—P and P≡P bonds are as 4 : 3 within the accuracy of the data.

TABLE 4

nature of bond	D (kX units)	D^2	remarks
O—O	1·46 to 1·47	2·13 to 2·16	see table 1
O=O	1·205	1·452	O_2 molecule, spectroscopic, r_e value
	Note: 3 × 1·452 = 4·356. 2 × 2·13 to 2·16 = 4·26 to 4·32		
N—N	1·47	2·16	see table 1
N=N	1·20 to 1·22	1·44 to 1·48	*
N≡N	1·0918	1·19	N_2 molecule, spectroscopic, r_e value
	Note: 2 × 2·16 = 4·32. 3 × 1·44 to 1·48 = 4·32 to 4·44. 4 × 1·19 = 4·76		
C—C	1·5412	2·3753	see table 1
C=C	1·337	1·79	C_2H_4, see p. 26
C≡C	1·201(6)	1·444	C_2H_2 spectroscopic
	Note: 3 × 2·3753 = 7·126. 4 × 1·79 = 7·16. 5 × 1·444 = 7·22		
P—P	2·18 to 2·21	4·75 to 4·88	see table 1
P≡P	1·8862	3·558	spectroscopic, r_e value
	Note: 3·558 ÷ 3 = 1·19 which is the D^2 value for N≡N. Also 3·558 × $\frac{4}{3}$ = 4·74 which is almost the D^2 value for P—P.		

* The values 1·20 to 1·22 are from different publications by L. Pauling.

For carbon the C≡C triple-bond length in acetylene is well established, and the single C—C bond length is known accurately for the diamond structure. The C=C double-bond length in ethylene is uncertain. The original value of 1·33 Å given by Penney (1937) has been shown to be too low, whilst the later value 1·3553 Å of Gallaway & Barker (1942) is now considered to be too high. We must acknowledge gratefully the help we have received on this point in discussion with Professor C. A. Coulson, according to whom the most probable value is 1·34 Å (= 1·337 kX). From table 4 it will be seen that the corresponding D^2 values for the C—C, C═C and C≡C bonds are almost in the ratio $\frac{1}{3}:\frac{1}{4}:\frac{1}{5}$. For the C═C and C≡C bonds the agreement with an exact whole-number ratio is within the accuracy with which the data are known for ethylene, whilst the deviation from a whole-number ratio for the C—C and C≡C values is of the order 1 part in 140 in D in spite of the great difference between the structure of the diamond and of a simple molecule. The line of approach adopted in the present paper is purely empirical, and we have therefore not discussed the effects of the theoretical 'bond orders' upon the D^2 values, but it may reasonably be claimed that the above figures suggest strongly that simple whole-number relations exist between the D^2 values for single and multiple bonds.

---

* For rubidium the data of Hume-Rothery & Lonsdale (1945) give the thermal expansion down to −183° C. For caesium no accurate data exist for either lattice spacings or coefficients of expansion and for the absolute zero we have used the value given by Simon & Vohsen (1928).

8. DISCUSSION

The relations described above are purely empirical and are probably contrary to what would be expected from the point of view of existing theory, and they are therefore presented with diffidence. It seems likely that the proportionality between D^2 and Z, which is shown in figure 1, possesses the greatest theoretical significance. According to modern views, what are usually called single covalent bonds may contain a small admixture of bonds of a higher order, and in view of the wide range of structures involved some of the points in figure 1 may require slight adjustment before they are exactly comparable. But even when full allowance is made for such complications it is difficult to avoid the conclusion that a proportionality between D^2 and Z is a real characteristic of the elements concerned in any one group of the first two short rows of the Periodic Table, and that this rule holds at least approximately when extended to the first long row.

It is more difficult to judge the significance of the results shown in figure 2, and much depends on the view that is taken of aluminium. If this element be regarded as normal, it will be seen that the points for boron, aluminium and scandium lie on a line which is concave downwards, and those for lithium, sodium and potassium on a curve concave upwards. It might therefore be suggested that the straight line through the points for beryllium, magnesium and calcium was fortuitous, although this would not explain the remarkable whole number sequence for the D^2 values. If, on the other hand, the abnormally low melting-point of aluminium indicates an abnormally weak binding, this may explain why the point for aluminium is higher than is required to give equal steps in D^2 on passing from Na → Mg → Al, and it would be less probable that the linearity of the points for beryllium, magnesium and calcium was merely fortuitous.

It might at first be thought that the diminution of D^2 by equal steps on passing from K → Ca → Sc, and from Rb → Sr → Yt was unlikely to indicate anything fundamental owing to the different co-ordination numbers of the alkali metal structures (body-centred cube, Co N. 8) and of the elements of groups II and III (Co N. 12). It must be remembered, however, that in the body-centred cubic structure each atom has not only its eight close neighbours at a distance equal to $\frac{1}{2}a\sqrt{3} = 0{\cdot}866a$, where a

is the side of the unit cube, but also six neighbours at a distance a which is only slightly greater. If we regard the valency electrons as giving rise to a direct bonding between neighbouring atoms, the bonding between second closest neighbours will be more important in the body-centred cubic than in the close-packed structures. Any direct bonding between second closest neighbours will reduce the fraction of the valency electrons available for bonding to the closest neighbours, and tentative calculations suggest that the body-centred and face-centred cubic structures may be more nearly comparable than is the case if only the closest neighbours are considered. The diminution of D^2 by equal steps on passing from $K \rightarrow Ca \rightarrow Sc$ and from $Rb \rightarrow Sr \rightarrow Yt$ may therefore indicate a general principle.

It is difficult to judge whether the various whole-number relations described above are of fundamental significance. The conditions in different structures will seldom be exactly comparable, so that if some underlying principle which involves whole numbers is really involved we shall in general expect an approximation to, rather than an exact agreement with whole number relations for the D^2 values. We may reasonably claim that the data given above show approximations to whole number relations which are more numerous and more exact than would be expected from pure chance, and we may suggest therefore that a theoretical investigation of the factors affecting the squares of the internuclear distances in crystals and molecules may reveal principles which have not yet been appreciated.

The author must express his gratitude to Professor Sir Cyril Hinshelwood, F.R.S., for laboratory accommodation and many other facilities which have greatly encouraged the present research. Thanks are also due to the Council of the Royal Society and to the British Non-Ferrous Metals Research Association for grants towards the cost of the work under the author's control.

REFERENCES

Gallaway, W. S. & Barker, E. F. 1942 *J. Chem. Phys.* **10**, 88.
Hume-Rothery, W. 1930 *Phil. Mag.* (7), **10**, 217.
Hume-Rothery, W. 1945 *Proc. Phys. Soc.* **57**, 209.
Hume-Rothery, W. & Lonsdale, K. 1945 *Phil. Mag.* (7), **36**, 842.
Levy, H. L. & Brockway, L. O. 1937 *J. Amer. Chem. Soc.* **59**, 2085.
Pauling, L. 1938 *Phys. Rev.* (2), **54**, 899.
Penney, W. G. 1937 *Proc. Roy. Soc.* A, **158**, 306.
Pitzner, K. S. 1945 *J. Amer. Chem. Soc.* **67**, 1126.
Simon, F. & Vohsen, E. 1928 *Z. phys. Chem.* (B), **133**, 180.

Theory of drafting of wool slivers. I

By D. R. Cox, *Wool Industries Research Association*

(*Communicated by H. Jeffreys, F.R.S.—Received* 19 *April* 1948—
Revised 1 *November* 1948)

The production of a textile yarn involves the drawing out of a thick intermediate web of fibres (sliver) by a process known as *drafting*. In the present paper, a mathematical study is made of the dynamical mechanism of drafting, with the object of investigating the effect of relevant factors on the irregularity of the yarn produced.

Equations are written down to govern the drafting of a very simplified model of a wool sliver, and a particular case, when the fibre length is small, is treated in detail. It is shown that when such a sliver is drafted by passing it between two pairs of rollers moving with different speeds, a 'period' is produced in the thickness variations of the drafted sliver. The change in standard deviation of thickness is related to the properties of the sliver and of the drafting mechanism, and in particular to the spectrum of the thickness of the initial sliver.

1. Introduction

(*a*) *Outline of problem*

One of the fundamental operations in the production of textile yarn from fibres is the process of *drafting*. It consists in the drawing out of a thick intermediate web of fibres, or sliver,* by one of two methods. In the first method, roller drafting, the sliver is passed between two pairs of rollers moving with different speeds, resulting in a reduction of thickness determined by the ratio of the roller speeds. The second method, mule drafting, is an intermittent process in which successive lengths of sliver are gripped, extended and then wound on a bobbin. The two methods of drafting are described more fully in § 1 (*b*).

It is important that the product of drafting should be as uniform as possible, because irregularities in the final yarn not only decrease its effective strength, leading to breakages in spinning and weaving, but may also spoil the appearance of the woven fabric. The sliver thickness after drafting depends on the laws of motion of what are known as the 'floating fibres'. For instance, in roller drafting, in which the distance between the rollers is usually arranged to be slightly greater than the length of the longest fibres, the short (floating) fibres are exposed during their passage between the rollers to the action of other fibres moving with different speeds. A similar effect is obtained in mule drafting, and the sliver thickness after drafting will in both cases depend on how the floating fibres have moved during drafting.

The aim of a theory of drafting is therefore to relate the changes of sliver thickness to the physical properties of the sliver and of the drafting mechanism, by relating them both to the statistics of fibre motion.

Now the thickness of a sliver, defined at each point of its axis by the number of fibres in the cross-section there,† is a typical example of a continuous time series.

---

* In practice a number of different terms are in use, corresponding to the stages of the intermediate product. For convenience the single term *sliver* will be used through this paper.

† Other definitions are possible in terms of, for example, the area occupied by fibres or the mass per unit length of the sliver. The definition given in the text will be used through the present paper.

A brief outline of the relevant parts of the theory of time series will be given in § 1 (*c*), but the application of the theory to such phenomena as Brownian motion, electrical noise and the fluctuations of economic series is now standard. The behaviour of sliver thickness may be summarized by the mean, the standard deviation and either by the autocorrelation function or by the normalized spectrum (these last two functions and the relation between them are discussed in § 1 (*c*)), and it is required to calculate the values of these statistical properties of thickness after drafting in terms of their values before drafting and the physical parameters of the sliver and the drafting machine.

Previous work on the changes of thickness in drafting has been experimental. The earliest work on the subject was by Balls (1928), and his conclusion, that in the roller drafting of cotton sliver a so-called 'drafting wave' is formed, has been confirmed and more deeply investigated in a series of papers by Foster and others, which has been summarized by Foster (1946). Spencer-Smith & Todd (1941) have reported similar phenomena with flax, while the whole position has been reviewed, with particular reference to wool, in a recent Physical Society Report by Martindale (1946). § 1 (*d*) of the present paper outlines some of Foster's conclusions about the drafting wave.

(*b*) *The process of drafting*

(i) *Introduction*

The present subsection gives a brief description of the two main types of drafting, roller drafting and mule drafting.

The methods used in drafting are similar in principle for all textiles, but the preparatory processes depend considerably on the type of fibre. A short account of all the stages in the production of yarn from wool fibres by the two distinct methods, worsted and woollen, is given in Martindale's (1946) Physical Society Report.

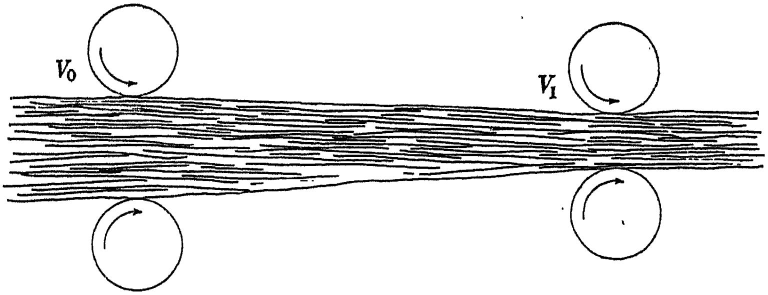


FIGURE 1. Roller drafting. Draft = V_1/V_0.

(ii) *Roller drafting*

In roller drafting, the sliver is passed between two pairs of rollers moving with different speeds (figure 1). The ratio of the surface speed of the front rollers to the surface speed of the back rollers, called the draft, is often between $1\frac{1}{2}$ and 10. The sliver is drawn out proportionally to the draft.

In the worsted process for the production of yarn, drafting starts from a sliver formed from large numbers of approximately parallel fibres of different lengths. This is passed through a series of drafting operations, and drafting is combined with *doubling*—the placing side by side of different lengths of drafted sliver to form a new thick sliver which is again drafted. Doubling leads, in general, to a more regular yarn.

In worsted roller drafting, the distance between the rollers is arranged to be slightly greater than the length of the longest fibre. Now if every fibre moved with the speed of the back rollers until it was actually gripped by the front rollers, the positions of the fibres after drafting would be known directly in terms of their positions before drafting. But in practice this does not happen. As the short fibres move between the rollers they are subjected to frictional forces due to contact with fibres moving with speeds between the back roller and front roller speeds, and the exact motion of the short fibres depends on the nature of these forces.

As the sliver gets thinner a small amount of twist is inserted to bind it together.

In the woollen process the fibres are much shorter and are not parallel, and the initial sliver is much thinner and is only drafted once.

(iii) *Mule drafting*

In the previous subsection an account has been given of roller drafting, and this is the more important method. Mule drafting is used in some forms of the woollen process, and it is of considerable theoretical interest as an alternative to roller drafting.

It has been seen that roller drafting is continuous. Mule drafting is intermittent; a length of sliver is gripped, extended and then wound on a bobbin. The sequence is then repeated with the next length of sliver. Twist is inserted during drafting.

Not all systems of mule spinning use mule drafting. For instance, although mule drafting is associated with mule spinning in the woollen process, roller drafting is used with mule spinning in the worsted process.

(c) *Specification of a continuous stationary time series*

(i) *General definitions*

It has been pointed out in § 1 (*a*) that the study of the variability of sliver thickness involves the analysis of a continuous stationary time series.

The theory of such series has been discussed by many authors, and so in the present paper the general concept of a stationary time series will be assumed known. It will, however, be convenient to outline the definitions of the normalized spectrum and the autocorrelation function, and to state the relation connecting them. A more detailed account has been given by Ming Chen Wang & Uhlenbeck (1945, p. 326). The theorems stated in this section will only be used in this paper in § 3.

Let $\{f(t)\}$ be a continuous stationary time series which may be supposed to have mean zero. In the present case $f(t)$ is the difference between the thickness at point t and the mean thickness. The autocovariance of $\{f(t)\}$ is defined by the equation

$$c_f(\tau) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(t)\,f(t+\tau)\,dt,$$

and the autocorrelation function $\rho_f(\tau)$ by

$$\rho_f(\tau) = \frac{c_f(\tau)}{c_f(0)}. \tag{1}$$

In the usual statistical language, $c_f(0)$ is the variance of $\{f(t)\}$, and $\rho_f(\tau)$ is the correlation coefficient between points τ apart.

Next assume the existence of

$$\mathbf{f}(p) = \lim_{T\to\infty} \frac{1}{(2T)^{\frac{1}{2}}} \int_{-T}^{T} f(t)\, e^{ipt}\, dt, \tag{2}$$

and let

$$\mathbf{F}(p) = \frac{|\,\mathbf{f}(p)\,|^2}{\left(\int_{-\infty}^{\infty} |\,\mathbf{f}(p)\,|^2 dp\right)}. \tag{3}$$

The function $|\,\mathbf{f}(p)\,|^2$ is called the spectrum of the time series and the function $\mathbf{F}(p)$ the normalized spectrum.

The autocovariance and the spectrum determine one another; in fact

$$c_f(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\,\mathbf{f}(p)\,|^2\, e^{-ip\tau}\, dp. \tag{4}$$

In particular the variance $c_f(0)$ is given by

$$c_f(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\,\mathbf{f}(p)\,|^2\, dp. \tag{5}$$

Therefore there is a similar relation to (4) holding between the autocorrelation function $\rho_f(\tau)$ and the normalized spectrum $\mathbf{F}(p)$. It is

$$\rho_f(\tau) = \int_{-\infty}^{\infty} \mathbf{F}(p)\, e^{-ip\tau}\, dp.$$

This and the corresponding inverse relation are known as the Wiener-Khintchine theorem. Since $\mathbf{F}(p)$ is even, the integral is usually written as a cosine transform, but the above form will be used in the present paper.

If the limit (2) defining $\mathbf{f}(p)$ does not exist, a more general form of the theory due to Wiener (1930) is required. The analysis given here is, however, sufficiently general for the present application.

The mean, variance and normalized spectrum (autocorrelation function) do not in general completely specify a time series, but they are the only properties that will be considered here.

(ii) *Application*

In dealing empirically with records of sliver thickness, it appears preferable to use the autocorrelation function rather than the normalized spectrum.

Foster (1946) showed that for cotton sliver the correlogram (graph of autocorrelation function) of sliver thickness after drafting is in many cases a damped harmonic curve of wave-length $2\pi/\nu$, say.

In such cases drafting is said to have produced a drafting *wave*. The value of the thickness of the sliver at the front roller at time* t has a maximum positive correlation with the value at time $(t + r\,2\pi/\nu)$ and a minimum negative correlation with the value at time $[t + \frac{1}{2}(2r+1)\,2\pi/\nu]$, where r is an integer. The use of the term *wave* in this way seems a reasonable statistical generalization of its use for an ordinary sine wave, where the maximum and minimum values of the correlation are plus and minus one. Any series whose correlogram tends to zero for large lags, and takes alternately positive and negative values, is of a similar type, and Yule (1926) suggested that the term *oscillatory* should be used in such cases.

The normalized spectrum $\mathbf{F}(p)$ is still of interest. If it has a maximum when p equals ν, then the correlation between $f(t)$ and a sine wave of wave-length $2\pi/p$, has a maximum when p equals ν, and so a period corresponding to this frequency will appear, on inspection, to show up above the random fluctuations.

As has been seen above, the normalized spectrum and the autocorrelation function are mathematically equivalent, and properly interpreted they give the same information. In the theoretical work in § 3 of this paper the discussion will be based on the normalized spectrum as it arises naturally.

(*d*) *Previous work on drafting*

A brief survey of the previous work on drafting has been given in § 1 (*a*). In this section a summary will be given of Foster's suggested explanation of the drafting wave.

It was explained in § 1 (*c*) that Foster found that in most cases the correlogram of sliver thickness after drafting is of the damped harmonic form. He gave empirical relations connecting the length of the drafting wave with the draft, twist and the distance between the rollers.

Foster suggested the following explanation of the drafting wave. Short fibres floating between the rollers are dragged forward by the more rapidly moving fibres already gripped by the front rollers. If a thick place occurs at the front rollers, more than the average number of fibres will be dragged forward. Thus a thick place passes through the front rollers leaving a thin place between the rollers. When this thin place reaches the front rollers, fewer floating fibres than on the average will be dragged forward leaving behind a thick portion. The process now repeats itself.

(*e*) *Scope of the present paper*

In the present paper the equations governing the drafting of a very idealized model of a sliver will be formulated.

In § 2 the basic physical assumptions are described and the equations written down. The physical assumptions have been chosen to apply to wool slivers. The drafting of slivers made of other fibres may require a different treatment at some points, although the same general method should be useful.

---

* This method of specifying the thickness of the sliver after roller drafting by giving the thickness at the front roller at a given time is more convenient than giving the thickness of the drafted sliver at a given position on its axis.

§ 3 is devoted to the detailed discussion of a very special case—the fibre length is supposed small compared with the distance between the rollers (roller drafting) or with the length drafted (mule drafting). This additional assumption is reasonably valid for drafting in the woollen process but not in the worsted process. The equations become partial differential equations which should apply to the drawing out of plastic substances not necessarily made of fibres.

2. PHYSICAL ASSUMPTIONS AND FORMULATION OF EQUATIONS

(*a*) *General remarks*

The equations governing the drafting of the idealized sliver will be formulated in § 6 (*c*), but first the physical assumptions will be outlined in § 6 (*b*). In § 6 (*d*) the assumptions are examined to see to what extent they agree with the properties of an actual sliver.

(*b*) *Statement of assumptions*

The fibre mass will be treated as negligible.

A reference axis will be taken along the axis of the sliver, and the position of a fibre will be specified by the co-ordinate of its right end and by its fibre chord—the difference in the co-ordinates of the right end and the left end. The fibre will be assumed to cross each section between its ends once only.

It will be postulated that the fibre chords remain unchanged during drafting. This would be so, for example, if the fibres were all parallel to the axis (or if the sliver had constant twist) and there was no elastic extension of the fibres in drafting.

The average coefficient of friction between a pair of fibres moving with relative velocity Δv is a function of Δv, $\phi[\Delta v]$, say. It will be assumed that the average coefficient of friction between fibres not in relative motion is given by the limiting friction $\phi[0]$. In some cases it will be necessary to work with a special form for $\phi[\Delta v]$, namely,

$$\phi[\Delta v] = \phi_1 \Delta v. \tag{6}$$

In some types of drafting oil is mixed with the fibres, and then the friction may be expected to be of the viscous type, when the law (6) would hold.

It will be assumed that in taking an average over all contacts at a given point between pairs of fibres of given chords, whose right ends are in specified positions,

$$\overline{\phi[\Delta v]} = \phi[\overline{\Delta v}]. \tag{7}$$

This implies that the function $\phi[\Delta v]$ is linear, or alternatively that all fibres in the same position have the same velocity.

The average normal pressure (force per unit length of contact) between two fibres in contact at x will be taken to be a function of the thickness at x, and of the time. For the thickness at a point may be assumed to determine the twist there, and the twist determines the compressive forces holding the fibres together.

It will be necessary to calculate the average number of contacts between z and $z+dz$ of fibres having their right ends in specified intervals $(w, w+dw)$ and $(y, y+dy)$. It will be assumed that the fibres at z occupy an area proportional to the total number of fibres in the cross-section there, and that the fibres from y and from w cross this area in points distributed randomly over the area.

Finally, the number of fibre contacts must be sufficiently large to justify the use of average coefficients of friction, average numbers of contacts, etc. The variations in coefficient of friction, in normal pressure, and in numbers of contacts will be assumed independent, so that the average of a product is the product of the averages of the separate factors.

(c) *Formulation of equations*

First of all consider a sliver with one end fixed ($x = 0$) and with a force $r(t)$ applied at the other end.

Let
$$n(l, x; t)\, dl\, dx$$
be the number of fibres which at time t have chord between l and $l+dl$, and right end between x and $x+dx$.

At time t the total number of fibres in the cross-section at x is

$$f(x,t) = \int_0^L dl \int_x^{x+l} n(l, y; t)\, dy, \tag{8}$$

where L is the length of the longest fibre.

Consider the set $S(x)$ of fibres with right ends to the right of x (see figure 2). There are three types of force acting on the fibres in the set $S(x)$:

(i) the applied force $r(t)$;

(ii) the friction between a fibre in $S(x)$ and a fibre not in $S(x)$, i.e. whose right end lies to the left of x;

(iii) the frictional forces between pairs of fibres both in $S(x)$.

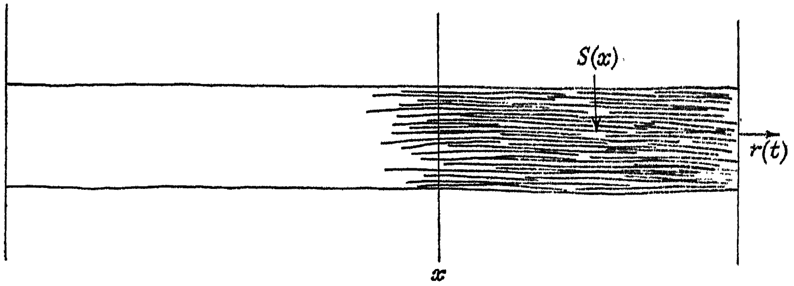


FIGURE 2. The set $S(x)$ of fibres.

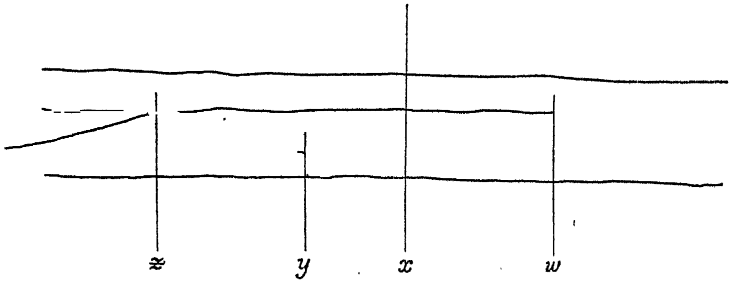


FIGURE 3. Notation for fibre contacts.

The forces making up (iii) balance out in pairs. Therefore since the fibre mass has been assumed negligible, the first two forces are in equilibrium.

It is now necessary to calculate the frictional force (ii).

Consider two fibres both intersecting the cross-section at z, and having right ends between w and $w+dw$ $(w>x)$, and y and $y+dy$ $(y<x)$ respectively (see figure 3).

The chance that they touch between z and $z+dz$ is

$$\alpha\, dz/f(z,\, t). \tag{9}$$

The average friction at such a contact is

$$g[f(z,t),t]\,\phi[\Delta v]\,\alpha\, dz/f(z,t), \tag{10}$$

where $g[f(z,t)\,t]$ is the average pressure between fibres in contact at z. The total force along the axis of the sliver developed at such contacts, for fibres of chords between l and $l+dl$, and h and $h+dh$ respectively, is

$$g[f(z,t),t]\,(\alpha\, dz/f(z,t))\,n(l,w;\,t)\,dl\,dw\times n(h,y;\,t)\,dh\,dy\,\phi[v(l,w;\,t)-v(h,y;\,t)], \tag{11}$$

where $v(l,x;\,t)$ is the average velocity at time t of fibres with chord l and right end at x.

The relation (7) has been used to express the average coefficient of friction in terms of the average relative velocity.

To get the total frictional force, the expression (11) is integrated over the following ranges:

$$\left.\begin{array}{lcccc} & y & \text{from} & z & \text{to} & [x,z+h],\\ & z & \text{from} & w-l & \text{to} & x,\\ & w & \text{from} & x & \text{to} & x+l,\\ \text{and} & l,h & \text{from} & 0 & \text{to} & L,\end{array}\right\} \tag{12}$$

where $[x,z+h]$ denotes the lesser of x and $z+h$.

Thus the equation of drafting is

$$\int_0^L dh\int_0^L dl\int_x^{x+l} dw\int_{w-l}^{x} dz\int_z^{[x,\,z+h]} dy\, g[f(z,t),t]\,\frac{\alpha}{f(z,t)}$$
$$\times\, n(l,w;\,t)\,n(h,y;\,t)\,\phi[v(l,w;\,t)-v(h,y;\,t)] = r(t). \tag{13}$$

With this is to be combined an equation of continuity

$$\frac{\partial n(l,x;t)}{\partial t}+\frac{\partial}{\partial x}[n(l,x;\,t)\,v(l,x;\,t)] = 0. \tag{14}$$

The functions n and v have been assumed to be differentiable.

In a more general treatment there would only be an equation of continuity for the total number of fibres with right ends between x and $x+dx$, since the fibre chords would change in drafting. Additional equations would then be required expressing the laws governing this change.

Equation (13) simplifies considerably if all the fibres have the same chord l. It becomes

$$\int_x^{x+l} dw\int_{w-l}^{x} dz\int_z^{x} dy\,\frac{g[f(z,t),t]}{f(z,t)}\,n(w,t)\,n(y,t)\,\phi[v(w,t)-v(y,t)] = r(t). \tag{15}$$

The equations have been obtained for a sliver gripped at $x = 0$, and with a force $r(t)$ applied at the other end. The equations are therefore immediately applicable to mule drafting, but they also apply to roller drafting, provided the force applied to the sliver at the front roller is uniformly distributed over the cross-section there. The end conditions to be satisfied in the two cases are, of course, different.

The problem is now this. A sliver whose initial thickness has given statistical properties (mean, variance and spectrum) is drafted in accordance with the above equations. What are the statistical properties of the thickness of the drafted sliver?

The mechanism of drafting as it has been formulated here is completely deterministic. That is to say the result of drafting a given sliver is always the same. Randomness only enters through the initial conditions.

(*d*) *Critical discussion*

In the previous subsections the equations governing the drafting of the idealized sliver have been derived and the underlying assumptions described. In this subsection the physical assumptions will be compared with the properties of actual slivers.

The method adopted consists in writing down the equation of motion of the set $S(x)$ of fibres whose right ends lie to the right of x. The inertial term is neglected—this is in general an excellent approximation. For example, the force required to draft a length of wool sliver of mass 10 g. is about 500 g. wt.

Experimental evidence will be produced elsewhere to show that wool slivers behave as if the coefficient of friction, $\phi[\Delta v]$, between two fibres increases almost linearly with their relative speed Δv. The explanation of this is not known yet and awaits a more detailed description of the motion of individual fibres in the sliver.

The most serious restriction on the analysis is the assumption that the friction between fibres not in relative motion can be calculated using the limiting friction $\phi[0]$. A more general treatment is desirable.

The equations have been developed to describe the drafting of slivers made from wool fibres. The same general methods should apply with other fibres.

To summarize, the assumptions on which the present paper is based are very simplified and require both extension and deeper analysis. The work does, however, show how such a very idealized sliver would behave when drafted.

3. DRAFTING WHEN FIBRE LENGTH IS SMALL

(*a*) *Introduction*

A very special case will now be treated in detail. It will be assumed that the fibre length is small compared with the length of sliver being drafted.

The results obtained will be directly applicable to drafting in the woollen process. They will not, however, apply to worsted roller drafting, where the distance between the rollers is arranged to be approximately equal to the length of the longest fibre.

In § 3 (*b*) the appropriate equations will be obtained from first principles, in §§ 3 (*c*) and 3 (*d*) mule drafting and roller drafting will be discussed, and the physical conclusions will be summarized in § 3 (*e*).

(*b*) *Drafting equations*

The equations of drafting could be derived by applying a limiting operation to the equations of § 2. Instead, they will be obtained directly from first principles.

As before let $f(x,t)$, called the thickness of the sliver, be the number of fibres in the cross-section of the sliver at position x and time t. Let $v(x,t)$ be the mean velocity there.

It will be assumed that the mechanism of drafting and the form of the initial sliver are such that $f(x,t)$ and $v(x,t)$ are differentiable functions of x and t.

Consider the equilibrium of those fibres with right ends to the right of x. The number of contacts with fibres with right ends to the left of x is, to a first approximation, $\alpha f(x,t)$, since $f(x,t)$ is continuous in x. The average relative velocity at the contacts is

$$\beta \frac{\partial v(x,t)}{\partial x},$$

where β depends on the fibre chord distribution near x, and will be assumed independent of x.

The equation corresponding to (13) is therefore

$$\alpha g[f(x,t),t]\, f(x,t)\, \phi\left[\beta \frac{\partial v(x,t)}{\partial x}\right] = r(t), \tag{16}$$

where $g[f,t]$ is the average normal pressure, $\phi[\Delta v]$ the coefficient of friction corresponding to a relative velocity Δv and $r(t)$ is the applied force.

The equation of continuity becomes

$$\frac{\partial f(x,t)}{\partial t} + \frac{\partial}{\partial x}[f(x,t)\, v(x,t)] = 0. \tag{17}$$

The equation (16) has been derived under the same conditions about the distribution of fibres as the more general equation (13). It is, however, valid under wider conditions. Suppose simply that the force acting on substance to the right of x due to substance to the left of x is the product of a function of the thickness at x with a function of the gradient of mean velocity at x. (This gradient measures the rate at which the substance is being pulled apart at this point.) Then an equation of the form (16) is obtained.

When the equations (16) and (17) are considered in this way, it seems probable that they will govern the drawing apart of a general plastic material if inertia is neglected.

(*c*) *Mule drafting*

In this section the solution of equations (16) and (17) will be obtained for end conditions corresponding to mule drafting.

Suppose that a length of sliver is gripped at $x = 0$ and that at the other end a force $r(t)$ is applied. The sliver will extend and its length at time t will be denoted by $a(t)$. It is required to find the thickness of the sliver at any time t.

The equations governing drafting are those that have been derived in the previous subsection. It should be observed that it has been assumed that the proportion of the frictional force contributed by static friction is negligible. This will only be the case if the sliver is drafting freely by fibre slippage.

It will be assumed that the variation in thickness is always small, so that $f(x, t)$, the thickness at position x and time t, can be written

$$f(x,t) = F(t)+\mu z(x,t), \tag{18}$$

where powers of μ above the first can be neglected, and where $F(t)$ is the mean thickness at time t.

Let the initial thickness of the sliver be given by

$$f(x,0) = F(0)+\mu z_0(x). \tag{19}$$

Now the velocity $v(x, t)$ corresponding to the drafting of a uniform sliver is $xV(t)$, where $a(t)\,V(t)$ is the velocity, $a'(t)$ of the free end.

On substituting into (16) and (17), it is found that

$$\alpha g[F(t),t]\,F(t)\,\phi[\beta V(t)] = r(t), \tag{20}$$

$$F'(t)+F(t)\,V(t) = 0. \tag{21}$$

The equation of continuity is equivalent to

$$F(t)\,a(t) = F(0)\,a(0). \tag{22}$$

Equation (22) can be obtained directly since the total number of fibres in the sliver is unchanged in drafting, or it can be deduced from (21) by using the condition at the free end

$$a(t)\,V(t) = a'(t). \tag{23}$$

When $F(t)$ is eliminated from (20) and (22), the load-extension curve of the uniform sliver is obtained.

Now consider the small perturbation in thickness given by (18). Let the corresponding perturbation in velocity when the force $r(t)$ is applied to the sliver be given by

$$v(x,t) = xV(t)+\mu y(x,t). \tag{24}$$

Substitute in equations (16) and (17), neglect terms in μ above the first, and use equations (20) and (21). Two partial differential equations for $y(x, t)$ and $z(x, t)$ are obtained. When $y(x, t)$ is eliminated between them, a Lagrangian equation is obtained for $z(x, t)$. The solution satisfying the initial condition (19) is

$$z(x,t) = \Theta(t)\,z_0\left[x\frac{F(t)}{F(0)}\right], \tag{25}$$

where
$$\Theta(t) = \frac{F(t)}{F(0)}\exp\left[\int_0^t \frac{\phi[\beta V(u)]}{\beta\phi'[\beta V(u)]}\left\{1+\frac{g_1[F(u),u]}{g[F(u),u]}F(u)\right\}du\right], \tag{26}$$

where
$$g_1[F(u),u] = \frac{\partial g[F(u),u]}{\partial F(u)}.$$

The ratio $F(0)/F(t)$ is the draft $D(t)$.

Thus it has been shown that if the initial thickness of the sliver is

$$f(x,0) = F(0)+\mu z_0(x),$$

the thickness at time t is

$$f(x,t) = \frac{F(0)}{D(t)} + \mu\Theta(t)\, z_0\left(\frac{x}{D(t)}\right). \tag{27}$$

This is the general result upon which the discussion of mule drafting will be based. The function $\Theta(t)$ depends on the behaviour of a uniform sliver when the force $r(t)$ is applied to it.

To interpret (27) consider again the special case of a uniform sliver. The velocity at a point x is

$$v(x,t) = xV(t), \tag{28}$$

and the mean motion of a point on the sliver initially at x_0 is governed by the equation

$$\frac{dx}{dt} = xV(t). \tag{29}$$

If equation (21) is used the position of the point at time t is found to be

$$x = D(t)\, x_0. \tag{30}$$

In the general case when the initial thickness is $f(x,0)$ equation (27) shows that $f(x,t) - F(t)$ is proportional to $f(x_0, 0) - F(0)$, x_0 being given by (30). Thus the irregularities of the sliver maintain their form during drafting and move along the sliver with velocity $xV(t)$. In other words, the whole pattern of thickness variation extends uniformly with the sliver. It should be noticed, however, that the motion of the actual fibres does not in general follow the pattern.

The equation (27) shows that the standard deviation of thickness of the length drafted has been multiplied by $\Theta(t)$ during drafting. The function $\Theta(t)$ depends on the nature of the sliver (i.e. on the functions ϕ and g) and also on the function $F(u)$ at all times between the beginning and end of drafting. Thus, in general, the change in standard deviation depends on the way the force $r(t)$ varies, and not only on the total draft. For instance, the change in standard deviation for drafting at constant force will depend on the value of the force, and for drafting at constant rate of extension will depend on the rate of extension.

Equation (26) can be used to find $\Theta(t)$ for any system; in what follows two further assumptions will be made which greatly simplify the calculation of $\Theta(t)$.

Suppose first that

$$\phi[\Delta v] = \phi_0 + \phi_1 \Delta v, \tag{31}$$

and secondly that

$$g[f(x,t), t] = g[f(x,t)], \tag{32}$$

or more generally

$$g[f(x,t), t] = g^{(1)}[f(x,t)]\, g^{(2)}[t].$$

The equation (32) has a simple physical interpretation. For a given sliver the value of g at a point is determined by the twist there, and this in turn is determined by the thickness at the point, and the average twist in the sliver. Now if a sliver is drafted and no additional twist is added, the average twist decreases as the length of the sliver increases. Thus in this case g is not simply a function of thickness. If, however, as drafting proceeds additional twist is inserted in such a way that the average twist is kept constant, then g is a function of thickness only (as expressed in equation (32)). This is what happens in practice.

With these two assumptions (31) and (32), it is found (after using equation (21)) that

$$\Theta(t)=\frac{g[F(0)]}{g[F(t)]}\exp\left[\frac{t\phi_0}{\beta\phi_1}+\int_0^t\frac{\phi_0 g'[F(u)]}{\phi_1\beta g[F(u)]}F(u)\,du\right]. \tag{33}$$

The equation (33) is valid for any $g[f]$. For the ratio of the final standard deviation to the initial standard deviation to be small, $g[f]$ should increase with decreasing f as rapidly as possible. In physical terms the twist should be as high as possible. Of course a practical limit is set by the introduction of fibre breakage.

When $\phi_0=0$, $\Theta(t)$ is equal to $g[F(0)]/g[F(t)]$ and is determined by the draft. The way the drafting force varies with time is irrelevant. When $\phi_0\neq 0$ this is not so, and this case must be investigated more closely.

Now the function $F(u)$ in equation (33) is the thickness at time u when the given force $r(u)$ is applied to a uniform sliver. If, however, as happens in practice, drafting is at constant rate of extension V_0, then, to a first approximation, $F(u)$ can be taken as the thickness at time u when the uniform sliver is extended at the constant rate V_0. The function $\Theta(t)$ can be evaluated for general $g[f]$, but it is convenient to make the additional assumption that

$$g[f]=\frac{Gf}{f-f^{(0)}}, \tag{34}$$

where G is a constant and $f^{(0)}$ is less than the least thickness of the sliver. This form of $g[f]$ has been chosen to show the effect of inserting twist in the sliver. Twist tends to run into the thin places increasing the pressure there, and to represent this, $g[f]$ increases with decreasing f.

It is found that

$$\Theta(t)=\left[\frac{F(0)-Df^{(0)}}{F(0)-f^{(0)}}\right]\exp\left[\frac{\phi_0 a(0)}{\phi_1\beta V_0}\left\{2D-2-\frac{F(0)}{f^{(0)}}\log\left(\frac{F(0)-Df^{(0)}}{F(0)-f^{(0)}}\right)\right\}\right]. \tag{35}$$

Now D exceeds unity and so

$$2D-2-\frac{F(0)}{f^{(0)}}\log\left(\frac{F(0)-Df^{(0)}}{F(0)-f^{(0)}}\right)>0. \tag{36}$$

Thus increasing V_0 and decreasing ϕ_0 decreases the final standard deviation, since in practice ϕ_0 is positive.

The main results of the present subsection can be summarized as follows. Equation (27) shows the result of the mule drafting of a sliver with general friction-velocity and pressure-thickness laws. To obtain this equation it has been assumed that the fibre length is short compared with the length drafted, and that the sliver drafts freely with only a negligible portion of the frictional force contributed by static friction.

The interpretation of equation (27) is that the whole pattern of thickness variation extends uniformly with the sliver. The change in the standard deviation of thickness of the length drafted depends on the friction-velocity and pressure-thickness laws, on the total draft, and on the way the drafting force varies with time. A special case, when the coefficient of friction is given by (31) and the pressure by (34), has been investigated in detail. When drafting is at constant rate of extension, V_0, the change in standard deviation of thickness depends on V_0 in the way shown by equation (35).

(*d*) *Roller drafting*

(i) *General remarks*

The solution of the equations for roller drafting will now be considered, the fibre length still being assumed small. In (ii) the mathematical theory is outlined, and in (iii) the physical significance of the results is discussed. In (iv) a very brief discussion is given showing how modifications of the drafting machine can be represented in the theory.

Two types of roller drafting will be considered here: (1) drafting at constant force, (2) drafting at constant roller velocity. The second is the practical case, but the investigation is simpler for the first type of drafting.

In drafting at constant force, the total force applied at the front roller is constant, and the velocity of the front roller varies, depending on the amount of material between the rollers. In drafting at constant roller velocity, the total force varies in such a way that the velocity of the front roller is constant.

Other types of drafting, in which the applied force and the roller speed are made to depend on the thickness of sliver between the rollers, are of practical importance and can be analysed by the methods of this section.

It will be assumed that the coefficient of friction $\phi[\Delta v] = \phi_1 \Delta v$ and the average normal pressure $g[f(x,t),t] = g[f(x,t)]$. These two forms have been discussed previously, the form for $\phi[\Delta v]$ corresponds to viscous flow, while the physical significance of the assumption about g has been discussed in the previous subsection (equation (32)). The form for g should always be valid in roller drafting.

(ii) *Mathematical theory*

The equations of drafting, (16) and (17), become

$$\alpha g[f(x,t)]\,f(x,t)\,\phi_1\beta\frac{\partial v(x,t)}{\partial x} = r(t),$$

$$\frac{\partial f(x,t)}{\partial t} + \frac{\partial}{\partial x}[f(x,t)\,v(x,t)] = 0. \tag{37}$$

The first equation will be rewritten as

$$g[f(x,t)]\,f(x,t)\frac{\partial v(x,t)}{\partial x} = r(t), \tag{38}$$

where the constant $(\alpha\phi_1\beta)$ has been absorbed into the function g.

A steady solution will be found for equations (37) and (38) corresponding to the drafting of a uniform sliver. A small perturbation from the steady state will then be considered.

To find the steady state, place

$$f(x,t) = F(x), \quad v(x,t) = V(x), \quad r(t) = R, \tag{39}$$

all independent of t. The equation of continuity (37) gives

$$F(x)\,V(x) = F_0 V_0, \tag{40}$$

where F_0 and V_0 are the values of $F(x)$ and $V(x)$ at the back roller ($x = 0$). Equation (38) gives

$$g[F(x)]\,F(x)\,V'(x) = R. \tag{41}$$

Introduce a small perturbation from the steady state solution, and write

$$v(x,t) = V(x)\,[1+\mu y(x,t)], \tag{42}$$

$$f(x,t) = F(x)\,[1+\mu z(x,t)], \tag{43}$$

$$r(t) = R[1+\mu w(t)], \tag{44}$$

where μ is small. Only the first order terms in μ will be retained.

Now
$$g[f(x,t)] = g[F(x)] + \mu F(x)\,z(x,t)\,g'[F(x)],$$

and so the equations (38) and (39) become

$$F(x)\frac{\partial z(x,t)}{\partial t} + V(x)\,F(x)\left[\frac{\partial y(x,t)}{\partial x} + \frac{\partial z(x,t)}{\partial x}\right] = 0, \tag{45}$$

$$y(x,t) + z(x,t) + z(x,t)\,F(x)\frac{g'[F(x)]}{g[F(x)]} - \frac{F(x)}{F'(x)}\frac{\partial y(x,t)}{\partial x} = -\frac{Rw(t)\,F(x)}{F_0 V_0 g[F(x)]\,F'(x)}. \tag{46}$$

To solve these equations it seems particularly appropriate to use the method of Fourier transforms, since it has been seen above that the properties of a time series are intimately connected with the Fourier transform. Suppose then that the mean values of y, z and w are zero and introduce the transforms $\mathrm{y}(x,p)$, $\mathrm{z}(x,p)$, $\mathrm{w}(p)$ as in § 1 (c), equation (2). Thus, for example,

$$\mathrm{y}(x,p) = \lim_{T\to\infty} \frac{1}{(2T)^{\frac{1}{2}}} \int_{-T}^{T} e^{ipt}\,y(x,t)\,dt.$$

If the above Fourier transformation is applied to the equations (45) and (46), two ordinary differential equations are obtained. The end conditions for these equations are:

$$\mathrm{y}(x,p) \equiv 0 \qquad \text{at} \quad x = 0; \tag{47}$$

$$\mathrm{z}(x,p) \equiv \mathrm{z}_0(p) \quad \text{at} \quad x = 0. \tag{48}$$

The first equation says that the velocity of the back roller is constant, while $\mathrm{z}_0(p)$ is the Fourier transform of the small variation of thickness fed in at the back roller.

A third condition is required and will depend on the type of drafting. For drafting at constant force

$$\mathrm{w}(p) \equiv 0; \tag{49}$$

for drafting at constant roller velocity

$$\mathrm{y}(x,p) \equiv 0 \quad \text{at} \quad x = a, \tag{50}$$

where a is the distance between the rollers.

Now $\mathbf{y}(x, p)$ can be eliminated between the two ordinary differential equations, and after some reduction the following equation is obtained for $\mathbf{z}(x, p)$:

$$\frac{d^2}{dx^2}\left[\mathbf{z}(x, p)\frac{F(x)}{F'(x)}\right] = \frac{ipF(x)}{F_0V_0}\frac{d}{dx}\left[\mathbf{z}(x, p)\frac{F(x)}{F'(x)}\right]. \tag{51}$$

From this the solutions for drafting at constant force and for drafting at constant velocity can be found. The value $\mathbf{z}_1(p)$ of $\mathbf{z}(x, p)$ at the front roller ($x = a$) is required.

For drafting at constant force, it is found that

$$\begin{aligned}\mathbf{z}_1^{(r)}(p) &= \mathbf{z}_0(p)\left\{\frac{F_0F_1'}{F_1F_0'} + \frac{F_1'}{F_1}\left(\frac{ipF_0}{V_0F_0'} - 1\right)I(p)\right\},\\ &= \mathbf{z}_0(p)\,\mathbf{m}^{(r)}(p), \quad \text{say.}\end{aligned} \tag{52}$$

For drafting at constant velocity

$$\begin{aligned}\mathbf{z}_1^{(v)}(p) &= \mathbf{z}_0(p)\left\{\frac{F_0F_1'}{F_1F_0'} + \frac{\dfrac{F_1'}{F_1}\left[1 + \dfrac{ip}{V_0}\left(\dfrac{F_1}{F_0'} - \dfrac{F_0}{F_0'}\right) - \dfrac{F_0F_1'}{F_1F_0'}\right]I(p)}{\exp\left[\displaystyle\int_0^a \frac{ipF(x)}{F_0V_0}\,dx\right] - 1 + I(p)\left(\dfrac{F_1'}{F_1} - \dfrac{ipF_1}{F_0V_0}\right)}\right\},\\ &= \mathbf{z}_0(p)\,\mathbf{m}^{(v)}(p), \quad \text{say,}\end{aligned} \tag{53}$$

where

$$I(p) = \int_0^a \exp\left\{\int_0^x \frac{ipF(y)}{F_0V_0}\,dy\right\}dx, \tag{54}$$

and

$$F_1' = F'(a).$$

The functions $|\,\mathbf{m}^{(r)}(p)\,|^2$ and $|\,\mathbf{m}^{(v)}(p)\,|^2$ will be called *spectrum multipliers.*

The discussion of roller drafting will be based on the two equations (67) and (68). It follows from the results stated in § 1 (*c*) that the ratio of the standardized variance of thickness after drafting to the standardized* variance before drafting is

$$\int_{-\infty}^{\infty} |\,\mathbf{z}_1(p)\,|^2\,dp \Big/ \int_{-\infty}^{\infty} |\,\mathbf{z}_0(p)\,|^2\,dp.$$

Also the autocorrelation function after drafting can if required be obtained by inverting the normalized spectrum. Thus the equations (52) and (53) give the statistical properties of the sliver thickness after drafting in terms of the properties before drafting.

One property of the spectrum multipliers $|\,\mathbf{m}^{(r)}(p)\,|^2$ and $|\,\mathbf{m}^{(v)}(p)\,|^2$ that will be required later, is that they both have the same limit as p tends to infinity. In fact

$$\lim_{p\to\infty} |\,\mathbf{m}^{(r)}(p)\,|^2 = \lim_{p\to\infty} |\,\mathbf{m}^{(v)}(p)\,|^2 = \left\{D\frac{g[F_0]}{g[F_1]}\right\}^2. \tag{55}$$

Explicit expressions for the actual thickness variation $z_1(t)$ can be found either by inverting (52) and (53) or directly by solving the partial differential equations (45) and (46) by the method of Monge. The final thickness is found to be a linear moving average of the initial thickness.

---

* The standardized variance is the variance divided by the square of the mean.

The equations (52) and (53) give solutions for a general pressure function $g[f]$. In the present paper the solutions for two particular forms of $g[f]$ will be considered in detail, one form to represent a twistless sliver and one to find the effect of the binding action of twist in thin places.

Now when the effect of twist can be neglected and the cross-sectional area of the sliver is proportional to the number of fibres in it, the average normal pressure at a point of the sliver between the fibres there will be independent of thickness:

$$g[f(x,t)] = G. \tag{56}$$

The steady state solution corresponding to the drafting of a uniform sliver is then

$$F(x) = F_0 \exp\left[-\frac{Rx}{GF_0V_0}\right],$$

and so the draft D is given in terms of the distance a between the rollers by

$$D = \exp\left[\frac{Ra}{GF_0V_0}\right]. \tag{57}$$

The general equations (52) and (53) express the spectrum multipliers in terms of an integral $I(p)$ given by (54). In the present case the equation for $I(p)$ gives

$$I(p) = AV_0 \mathrm{e}^{iAp} E(Ap), \tag{58}$$

where

$$A = GF_0/R \tag{59}$$

and

$$E(x) = \int_{x/D}^{x} (\mathrm{e}^{-iu}/u)\, du. \tag{60}$$

The spectrum multipliers can now be evaluated, and it is found that for drafting at constant force

$$\mathbf{z}_1^{(r)}(p) = \mathbf{z}_0(p)\, \mathbf{m}^{(r)}(p),$$

where

$$\mathbf{m}^{(r)}(p) = 1 + (1 + iAp)\, \mathrm{e}^{iAp} E(Ap), \tag{61}$$

and for drafting at constant velocity

$$\mathbf{z}_1^{(v)}(p) = \mathbf{z}_0(p)\, \mathbf{m}^{(v)}(p),$$

where

$$\mathbf{m}^{(v)}(p) = \frac{E(Ap)\,(1 + iAp) + \mathrm{e}^{-iAp} - \mathrm{e}^{-iAp/D}}{E(Ap)\,(1 + iAp/D) + \mathrm{e}^{-iAp} - \mathrm{e}^{-iAp/D}}. \tag{62}$$

The real and imaginary parts of $E(Ap)$ are respectively the difference of two cosine integrals and the difference of two sine integrals. Thus $E(Ap)$ can be found from standard tables, and so the final spectra, $|\mathbf{z}_1^{(r)}(p)|^2$ and $|\mathbf{z}_1^{(v)}(p)|^2$, can be evaluated in terms of the initial spectrum, $|\mathbf{z}_0(p)|^2$, and the constants D and A.

The function $|\mathbf{z}_0(p)|^2$ characterizes the initial sliver, and the constants D and A characterize the drafting machine. The constant D is the draft, and so it remains to give the physical interpretation of A. The formula (58) for A can be rewritten, using the expression (57) for the draft,

$$A = a/(V_0 \log D). \tag{63}$$

The dimensions of A are those of time, and for fixed draft and back roller velocity, A is a measure of the distance between the rollers. It can be shown that the mean time taken by a fibre to pass between the rollers is $(D-1)A/D$.

The analysis is very similar when the effect of twist is represented by the pressure law

$$g[f] = \frac{Gf}{f-f^{(0)}}. \tag{64}$$

The solution is expressed in terms of three parameters. The first, k, characterizes the twist in the sliver,

$$k = f^{(0)}/F_0 \quad (0 < k < 1/D), \tag{65}$$

the second is the draft D, and the third, B, has the dimensions of time, and is such that

$$-B \log\left\{\frac{(1-Dk)}{D(1-k)}\right\} \tag{66}$$

is equal to the time that a fibre would take to move between the rollers if it continued to move with the speed of the back rollers.

The explicit expressions for the spectrum multipliers can be written down from the general relations (52) and (53). The detailed expressions will not be given here as they are cumbersome. Some numerical results are reported in the next subsection.

(iii) *Discussion of solution*

The mathematical discussion of roller drafting has been given in § 3 (*d*) (ii). In the present subsection special cases will be discussed numerically, and the physical significance of the conclusions considered.

The case when the initial thickness variations are not autocorrelated (i.e. are completely random)

$$\rho_0(\tau) = \begin{matrix} 1 & \tau = 0; \\ 0 & \tau \neq 0, \end{matrix} \tag{67}$$

will be considered first. This form for $\rho_0(\tau)$ can only be approached as a limiting case, since for the thickness to be continuous and differentiable there must be high correlation between the thickness at points a short distance apart (see, for example, Slutzky (1928)). However, the form (67) can be used if the autocorrelation function is nearly unity for lags τ comparable to the fibre length, and nearly zero for lengths comparable to the distance between the rollers.

It can be deduced from the result that the spectrum multipliers tend to the constant value (55) at infinity, that the thickness after drafting still has autocorrelation function (67), and that the ratio of the coefficient of variation of thickness after drafting to the coefficient of variation before drafting is

$$Dg[F_0]/g[F_0/D], \tag{68}$$

where F_0 is the initial mean thickness. In the special case of a twistless sliver in which the pressure function $g[f]$ is constant, the effect of drafting is to multiply the coefficient of variation by the draft D. When the effect of the insertion of twist is represented by the pressure function (64)

$$g[f] = Gf/(f-f^{(0)}),$$

where

$$f^{(0)}/F_0 = k,$$

the coefficient of variation is multiplied by $D(1-Dk)/(1-k)$, and since D exceeds unity, this decreases with increasing k (i.e. increasing twist), as would be expected on physical grounds.

It should be emphasized that the above results only hold when the initial thickness variations have zero autocorrelation, as given by (67). When the initial autocorrelation function of thickness is not of this form it is necessary to find the change of the coefficient of variation and of the spectrum numerically, by determining the spectrum multipliers from the formulae given above ((52) and (53)).

This has been done for drafting without twist with

$$a/V_0 = 2\log 2, \quad D = 2 \text{ and } 8. \tag{69}$$

The notation has been explained in the previous subsection. The time a fibre would take to move between the rollers if it continued to move with the speed of the back rollers is a/V_0, and this, together with the draft D, characterize the drafting machine. The effect of the insertion of twist has been examined by taking the pressure law

$$g[f] = Gf/(f-kF_0)$$

with

$$k = 0{\cdot}4,$$

and

$$a/V_0 = 2\log 2, \quad D = 2.$$

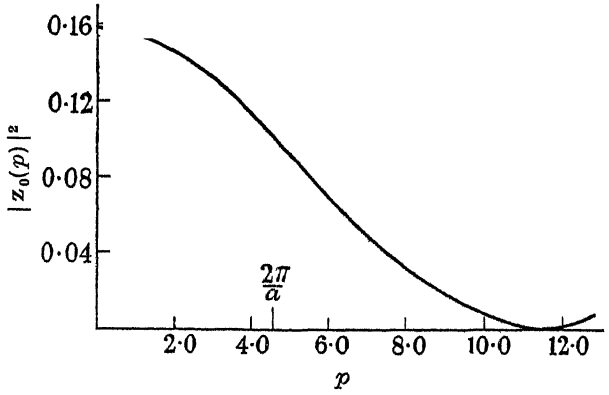


FIGURE 4. Initial normalized spectrum corresponding to linear moving average of extent one-half.

The choice of the initial spectrum is at this stage arbitrary. It can, however, be seen that the general form of the spectrum after drafting does not depend critically on the exact form of the initial spectrum, and that the general conclusions set out below are deduced from the form of the spectrum multipliers, and will hold for a broad class of initial spectra. The exact initial spectrum used has been chosen to correspond to a linear moving average of extent one-half:

$$\rho_0(\tau) = \begin{matrix} 1-2|\tau|, & 0 \leqslant |\tau| \leqslant \frac{1}{2}; \\ 0, & \frac{1}{2} \leqslant |\tau|; \end{matrix} \tag{70}$$

$$|z_0(p)|^2 = \frac{4}{\pi p^2}[1-\cos(p/2)]. \tag{71}$$

The initial spectrum is plotted in figure 4 and the spectrum multipliers and the spectra after drafting in figures 5 to 7.

The thickness of the sliver before and after drafting has been specified by the thicknesses at the back and front roller at a given time. Thus the argument p of the spectra is of the form 2π/wave-length, where the wave-length is measured in time.

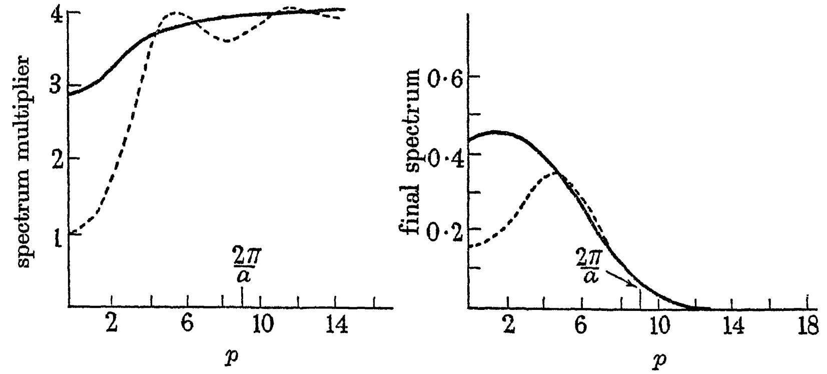


FIGURE 5. Spectrum multipliers and final spectra for drafting with $g[f]$ = constant, $\phi[\Delta v] = \phi_1 \Delta v$, and draft two. ——— drafting at constant force. ------ drafting at constant velocity.

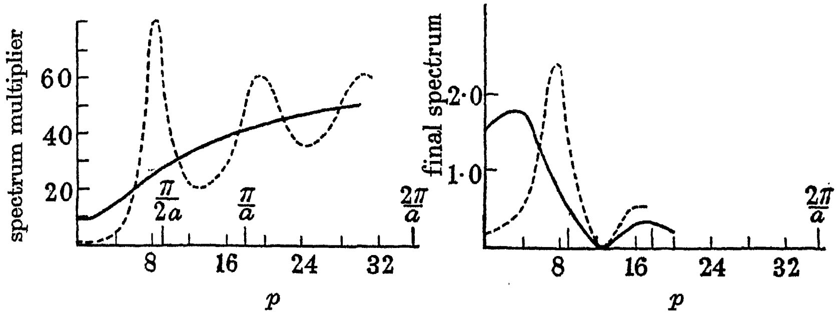


FIGURE 6. Spectrum multipliers and final spectra for drafting with $g[f]$ = constant, $\phi[\Delta v] = \phi_1 \Delta v$, and draft eight. ——— drafting at constant force. ------ drafting at constant velocity.

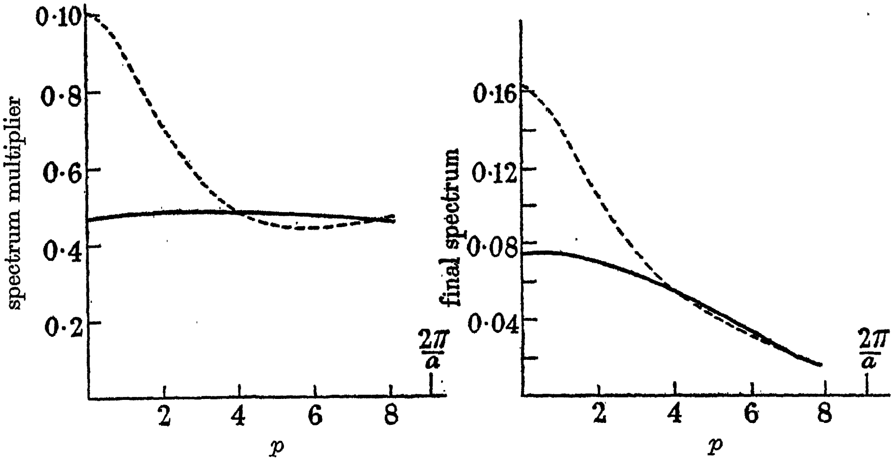


FIGURE 7. Spectrum multipliers and final spectra for drafting with $g[f] = Gf/(f-f^{(0)})$, $\phi[\Delta v] = \phi_1 \Delta v$ and draft two. ——— drafting at constant force. ------ drafting at constant velocity.

To convert into the form 2π/wave-length, where the wave-length is measured along the sliver in units of the distance, a, between the rollers, divide the value of p by $2\log 2$ for the initial sliver, and by $2D\log 2$ for the final sliver. The value of p corresponding to 2π/(distance between the rollers) is marked on all the figures.

The initial spectrum has been normalized and the final spectrum has not been normalized, so that the area (from zero to infinity) under the final spectrum gives the ratio

$$\left\{\frac{\text{coefficient of variation after drafting}}{\text{coefficient of variation before drafting}}\right\}^2.$$

The main conclusion for drafting with $g[f]$ constant is that when drafting is at constant roller velocity, a drafting wave is formed for a broad class of initial spectra. The drafting wave is very pronounced at the higher draft (eight), and in this case there are at least three prominent peaks in the spectrum multiplier, which will lead, for most initial spectra, to peaks in the final spectrum. At the lower draft (two), the tendency to oscillate is less. However, for many initial spectra which steadily decrease for values of p between zero and ten, a single peak is formed in the final spectrum.

A drafting wave is much less likely to be formed when drafting is at constant force, but for most initial spectra the coefficient of variation of thickness is greater after drafting at constant force than after drafting at constant roller velocity. This is very reasonable physically. For in drafting at constant force a thick portion of sliver between the rollers reduces the roller speed, leading to a reduced draft. If the initial thickness variations contain a period corresponding to the peaks in the spectrum multiplier, the above conclusion is not true, and drafting at constant roller velocity leads to a large increase in the coefficient of variation of thickness.

The pressure-thickness law $g[f] = Gf/(f-f^{(0)})$ which has been taken to simulate the effect of twist, appears both to reduce the coefficient of variation of thickness and to suppress the tendency to oscillate. This suppression of the drafting wave is not in accord with Foster's experimental conclusion that in the roller drafting of long fibres twist accentuates the drafting wave.

The coefficient of variation of sliver thickness is less after drafting at constant force than after drafting at constant roller velocity in the special case considered.

(iv) *Two generalizations*

It will be convenient to mention briefly at this point two generalizations of the work on roller drafting. Both refer to attempts to reduce the increase in irregularity of thickness that occurs in drafting.

In the Casablanca system of drafting, the normal pressure is varied along the axis of the sliver. The aim is to increase the control over the fibres near the front roller. This device can easily be represented in the general drafting equations by replacing the pressure function $g[f,t]$ by $g[f,t]\,A(x)$. In the particular case when the fibres are all short, equation (38) is generalized to

$$A(x)\,g[f(x,t)]\,f(x,t)\frac{\partial v(x,t)}{\partial x} = r(t). \tag{72}$$

When $A(x) = e^{kx}$ a solution for the spectrum multiplier can be obtained by the methods of § 3 (*c*).

Another method of attempting to reduce the thickness variations is to vary the force applied at the front rollers (and so the front roller velocity) in a way depending on the thickness of the sliver fed in at the back roller. The idea of doing this appears to be well known in the textile industry, although I have been unable to find published work on the subject. If the notation of § 3 (*c*) is used, $\mu w(t)$ is the deviation of applied force and $\mu z_0(t)$ is the deviation of initial thickness, both measured from their mean values. The most general linear mechanism of the type being considered would make the deviation in applied force a function of the initial thickness, past, present and future, and so would give

$$w(t) = \int_{-\infty}^{\infty} b(u)\, z_0(t-u)\, du, \tag{73}$$

where $b(u)$ is a weighting function. If the Fourier transformation considered previously is applied to (73) the simple form

$$\mathbf{w}(p) = \mathbf{b}(p)\, \mathbf{z}_0(p)$$

is obtained, where

$$\mathbf{b}(p) = \int_{-\infty}^{\infty} e^{ipu}\, b(u)\, du.$$

The spectrum multiplier $|\mathbf{m}^{(b)}(p)|^2$, corresponding to the weighting function $b(u)$, can be found by following the method given in § 3 (*c*). A detailed discussion will not be given here, since a device of this form is only likely to be used in drafting long fibres. There is, however, one point of immediate interest. It can be shown that provided

$$|\mathbf{b}(p)| = o(p) \quad \text{as} \quad p \to \infty, \tag{74}$$

$\lim_{p\to\infty} |\mathbf{m}^{(b)}(p)|^2$ is independent of $b(u)$. The condition (74) is likely to be satisfied for all $b(u)$ of practical interest—it will be satisfied if $b(u)$ contains Dirac delta functions, but not if it contains the derivatives of such functions. (In physical terms $w(t)$ may depend on the value of the thickness at isolated points, but not on the rate of change of thickness.)

Now the final variance depends on the integral

$$\int_{-\infty}^{\infty} |\mathbf{z}_0(p)|^2\, |\mathbf{m}^{(b)}(p)|^2\, dp,$$

and for given initial spectrum can only be reduced in size by reducing $|\mathbf{m}^{(b)}(p)|^2$. The above result shows that the type of mechanism considered can only reduce $|\mathbf{m}^{(b)}(p)|^2$ for small values of p. Now for the smaller drafts at least it is the rise of $|\mathbf{m}(p)|^2$ from small p to large p that produces the drafting wave. It is therefore possible that in a mechanism of this form reduced variance will be accompanied by an increased tendency to give a drafting wave.

The results of (iii) comparing drafting at constant force with drafting at constant roller velocity exemplify this conclusion. Drafting at constant velocity leads to reduced variance and to an increased tendency to give a drafting wave.

The account of this section is very brief because these modifications are only used in drafting long fibres. It is, however, very important to see how these modifications can be represented mathematically because one of the main practical applications of the theory is likely to be to new drafting machines.

(e) *Summary*

A simplified model of a sliver has been considered in which the fibre length is supposed small. This assumption is approximately valid in the woollen process where the length of sliver drafted considerably exceeds the fibre length.

It is shown that in the mule drafting of such a sliver the final coefficient of variation of the length drafted depends on the law connecting friction with velocity, the law connecting the pressure between fibres with sliver thickness, and on the way the drafting force is applied. Where the coefficient of friction is a linear function of velocity increasing with increasing velocity, the final coefficient of variation for given draft is smaller if the drafting speed is high. In the general case no drafting wave is obtained and the profile of the sliver is unchanged in drafting.

Roller drafting has been investigated with the additional assumptions that (i) the average pressure between the fibres in contact at a given point is independent of the thickness there; (ii) the coefficient of friction between two fibres in contact is, on the average, proportional to their relative velocity. It is shown that a drafting wave is generally formed when such a sliver is drafted at constant roller velocity, although whether or not a drafting wave is found in any particular case depends on the initial spectrum of thickness. The tendency to form a drafting wave appears to grow more pronounced with increasing draft. A drafting wave is much less likely if the applied drafting force is constant, but in this case the coefficient of variation of thickness after drafting is, for most initial spectra, greater than for drafting at constant roller velocity.

When the effect of the binding action of twist in thin places is represented by the special pressure-thickness law

$$g[f] = Gf/(f-f^{(0)}),$$

the drafting wave is suppressed and the coefficient of variation of thickness after drafting reduced.

If the initial thickness is completely random—a limiting case only—the effect of drafting is to multiply the coefficient of variation of thickness by

$$Dg[F_0]/g[F_0/D],$$

where D is the draft, F_0 the initial mean thickness, and $g[f]$ is the average normal pressure between fibres at a point where the sliver thickness is f. The autocorrelation function is unaltered.

4. FURTHER WORK

The present paper has given drafting equations for a very idealized model of a sliver, with the detailed solution only for the special case when the fibre length is small. In a later paper it is hoped to deal in detail with drafting when the fibre length is not small, and also to give a deeper analysis of the principles underlying the motion of the fibres. The mathematical model can also be used to investigate various modifications of conventional drafting machines.

The problem was suggested by Dr H. E. Daniels and the research began in close collaboration with him. I wish to express my deep indebtedness to Dr Daniels for this collaboration during a period in which a number of the basic results of the paper were developed. Dr Daniels has also made many valuable suggestions at later stages of the work.

I wish to thank Mr B. H. Wilsdon, Director of Research, Wool Industries Research Association, for his interest in the work and for permission to publish. Mr S. L. Anderson has been responsible for the experimental work connected with this problem and I have had the advantage of many discussions with him on the physical aspects of drafting.

References

Balls, W. L. 1928 *Studies of quality in cotton.* London: Macmillan.
Foster, G. A. R. 1946 *J. Roy. Statist. Soc. Suppl.* **8**, 42.
Martindale, J. G. 1946 *Rep. Progr. Phys.* **10**, 172.
Ming Chen Wang & Uhlenbeck, G. E. 1945 *Rev. Mod. Phys.* **17**, 323.
Slutzky, E. 1928 *C.R. Acad. Sci., Paris*, **187**, 878.
Spencer-Smith, J. L. & Todd, H. A. C. 1941 *J. Roy. Statist. Soc. Suppl.* **7**, 131.
Wiener, N. 1930 *Acta Math.* **55**, 117.
Yule, G. U. 1926 *J. Roy. Statist. Soc.* **89**, 1.

Explosion waves and shock waves. VII. The velocity of detonation in cast T.N.T.

By W. B. Cybulski, (the late) W. Payman and D. W. Woodhead

(*Communicated by Sir Robert Robertson, F.R.S.—Received* 3 *August* 1948—*Revised* 12 *November* 1948)

[Plates 2 to 4]

The Buxton rotating-mirror camera has been used for the precise determination of velocities of detonation in cylinders of cast T.N.T. The method has an advantage over the indirect Dautriche method in that it affords a direct measure of the velocity. The effects on the velocity of detonation of the following variables have been studied in so far as each can be independently varied: (*a*) the purity and crystal size, (*b*) the diameter and (*c*) the degree of confinement. Correlation of the observations with those of the velocity of detonation shows that cast T.N.T. may undergo stable or unstable detonation according as the crystals are small or large respectively. The unstable regime is characterized by the failure of the detonation wave to be propagated uniformly over the full cross-section of the cylinder of explosive.

The velocity of detonation of T.N.T. rises as the diameter of the cylinder is increased from a 'boundary diameter', below which there is no propagation, to a 'limiting diameter' at which the velocity reaches a maximum value. In castings of density 1·625 the maximum velocity is 6950 m./sec. The value of the boundary diameter is a measure of the sensitivity to propagation. The effect of heavy confinement is to reduce the boundary and limiting diameters; the maximum velocity, however, remains unaltered and confinement has therefore the same effect as an increase in diameter in an unconfined charge.

INTRODUCTION

Factors affecting the velocity of detonation

The velocity of detonation of a solid explosive varies within limits with changes in its density, purity, crystal structure, size, shape and degree of confinement. The general effects of some of these factors have been known for some time (Abel 1873–4; Berthelot 1885; Dautriche 1906), and the present experiments have been made to examine them in greater detail and with a higher accuracy than has been achieved hitherto. In the theoretical consideration of velocities of detonation it is important to know whether or not the velocity measured is that of a stable detonation. A photographic method was chosen in order to obtain a continuous record showing the stability of the detonation in addition to providing a measure of velocity. (The usual criterion of stable detonation is that its velocity should be constant along successive lengths of the cartridge. Hence, a degree of both persistence and regularity in propagation is stipulated. No method other than the photographic can provide this information.)

Many factors combine to control the velocity of detonation of an explosive, and it is not possible always to isolate one for particular study. For example, similarity of both crystal size and density cannot be ensured with castings of different diameters. The separate effects of the following variables on the velocity of detonation of castings of T.N.T. have been examined in so far as each of them can be independently varied: (*a*) the physical state of the explosive in respect of its purity and crystal size (with special reference also to cast amatols), (*b*) the diameter of the cylindrical charge, and (*c*) confinement.

The effects of the diameter of the charge and of the material and thickness of the confining tube are of theoretical interest in that they provide a measure of the influence of surface conditions on the velocity of detonation in charges of small diameter. With charges either of sufficiently large diameter or under adequate confinement, the effect of lateral expansion of the products during the reaction time can be neglected, and within the charge the energy is directed wholly to the maintenance of the detonation wave. Under these conditions the velocity of detonation reaches a maximum.

Photographic records of unconfined charges

Figure 1, plate 2, is a photographic record of the detonation of a cylindrical charge of crystalline T.N.T., packed to a density of 1·00 in a cellophane tube; the cylinder was 3·17 cm. (1¼ in.) in diameter and 30 cm. (12 in.) long. The trace of the detonation wave is smooth and straight, indicating a uniform velocity throughout (4650 m./sec.), and it is followed continuously by the record of luminous products expanding laterally; the general slope does not indicate the velocity of the lateral flame, since its angle of projection is not known, but the length gives some idea of the duration of burning. As pointed out by Jones (1928, p. 617) the thickness of the trace of the detonation wave may be a guide to the thickness of the detonation or reaction zone; this can be derived simply from measurements of the trace and the known width of the camera slit, and for the record in figure 1 it is found to be 2·6 mm. (0·55 μsec.).

The reliability of such measurements depends on two assumptions. The first is that the zone of luminosity is able to affect the whole of the corresponding area of the film, but this is unlikely and therefore measured values of the length of the zone may be too short. The second assumption is that the thickness of the recorded trace is not modified by forward motion of the luminous elements during the period of reaction; if such translational motion does occur, the value of the zone length must again be too short. Further uncertainty may be introduced if the products remain incandescent after the reaction is complete.

Figure 2, plate 2, was obtained with a charge of T.N.T. cast 'creamed', that is, of fine crystal structure; the diameter was 3·17 cm., length 40 cm. and density 1·62. The record differs from that in figure 1 in that the trace of the wave front is much narrower and is separated from that of the lateral flame by a fine dark band; owing to the higher speed of the wave (6900 m./sec.), the record is more nearly normal to the direction of motion of the film. The thickness of the reaction zone in this charge is assessed as only a small fraction of a millimetre, but this value is probably reduced by the effects mentioned above. It is evident that in both figures 1 and 2 the detonation is regular and stable.

It is equally evident that an unstable regime of detonation obtains in figure 3, plate 2, which is a record of the wave in a cast T.N.T. of coarse crystal structure; the dimensions of the charge were the same as that in figure 2. With such records the velocity of detonation is measured over any interval which shows a reasonably uniform slope. The appearance of the luminosity behind the wave front is intermittent, and it is suggested that this saw-tooth effect is an indication of instability, even when the velocity of detonation is uniform over an appreciable length of explosive. The dark areas in the record immediately behind the wave front are very similar to those formed when detonation is dying out (as it may) in a charge of cast T.N.T. of coarse crystal structure or in a charge of small diameter; they may be due to the fact that some explosive in the surface layers escapes full decomposition, the detonation wave propagating through a core of explosive, weakly confined by the undecomposed layers. The diameter of the core either shrinks until 'fading' of the detonation occurs or it grows again to that of the cylinder and the detonation wave is completely re-established, a sequence which, if repeated, gives rise to a record with the irregular saw-tooth appearance of figure 3. The front of the submerged wave, seen as a thin line joining the heads of the dark spaces, is photographed because of the translucency of the cast explosive, for if the surface is blackened the trace disappears until such time as a detonation wave extends to the surface again. The slopes of the faint lines at the head of the dark spaces are variable and correspond with velocities as low as 5000 m./sec., values which may be characteristic of the diameter of the core at any instant. The above may not be the full explanation of the saw-tooth effect, since there is one contrary observation. On some photographic records taken with the slit at right angles to the axis and just beyond the end of the cylinder, there is a feebly luminous disk, fully as wide as the charge at the head of the disturbance; this is shown in figure 4, plate 2, which is an enlarged print of the negative. The disk is followed immediately by a brighter zone, conical in shape and narrower than the diameter. There is a pause then before the appearance of the lateral flame. The corresponding record, figure 5, plate 2, of the detonation of a charge of fine-grained T.N.T. (creamed) shows a disk of intense flame as wide as the cartridge, followed immediately by the spreading lateral flame, and it may well be that at the particular instant shown in figure 4 with the coarse-grained casting (poured-clear) the detonation wave was about to be re-established over the whole cross-section.

Photographic records of confined charges

A record obtained with creamed T.N.T. cast in a glass tube, of internal diameter 3·17 cm. and wall thickness 1 mm., is reproduced in figure 6, plate 3. The dark space between the traces of the wave and the luminous effects spreading laterally is due to the obscuration of light by glass dust, the tubing being pulverized by the passage of the detonation wave; since the glass is of uniform thickness the dark band has a constant width. Figure 7, plate 3, was obtained with a coarse T.N.T. cast in a similar glass tube; the rate of propagation of the explosion is far from uniform, and here the dark band arises from the effects both of shattered glass and of annular undetonated explosive as previously described. The value of the photographic method is emphasized by this record, which shows that stable detonation is not always obtained, even though the mean rate over short distances may be sensibly uniform and reproducible from round to round.

A record similar to figure 6 is reproduced in figure 8, plate 3, which is of an unconfined cast charge of a 50:50-amatol; the trace of the wave is straight, but the charge does not appear to have detonated over the whole of its cross-section, the surface layers escaping to give rise to the dark space between the wave and the following luminous gases.

To photograph a detonation within an opaque tube it is necessary to provide windows without unduly weakening the material; the method we have applied has been to drill a line of small holes in the wall of the tube. That the consequent loss of confinement is insufficient to cause a reduction in the velocity has been shown by measurement with tubes in which the holes were spaced, first at different intervals, secondly over a short distance near each end with the central portion of the tube left whole; there were no detectable differences in over-all velocity. The holes, each 2 mm. in diameter and spaced at 1 cm. intervals, were plugged with lead wire during the filling process. The definition of a record is improved by sealing the inner ends of the holes with a strip of transparent adhesive tape before casting. A 28 g. primer pellet of pressed tetryl was not enclosed in the steel tube. A typical record is shown in figure 9, plate 3; the explosive was a fine T.N.T. cast into a steel tube, 3·17 cm. internal diameter, 45 cm. long and 3·2 mm. wall thickness. The record is sufficiently sharp for accurate measurement; the streaks below the trace of the wave itself are due to flame projected through the holes and give no indication of what is happening behind the wave within the tube which is rapidly fragmented. Figures 10 and 11, plate 3, differ from figure 9 only in that the holes are spaced 1·5 and 2 cm. apart respectively; all gave the same velocities. The intensity of the record does not seem to depend on the diameter of the charge, as it does when unconfined, for equally good records are obtained with confined charges of only 6 mm. diameter; it is probable therefore that it is flame in the holes rather than the detonation wave itself which is photographed. However, the slope of the line passing through the extreme tips of the individual traces can be accepted as indicating accurately the rate of the detonation wave itself. The camera must have a clear view to the bottom of each hole and, in a long tube with a thick wall, it is necessary to mill a channel along the line of the holes, or to countersink each one; neither of these modifications has any measurable effect on the speed of the wave.

Criteria of stable detonation

The criteria of stable detonation are usually considered to be threefold: that the detonation proceeds as a shock wave at a steady rate over an appreciable length of explosive, that the wave is maintained by the energy liberated in the shock front, and that the explosive ahead of the wave remains completely undisturbed. Indirect methods of measurement such as the Dautriche (1906) can give information only in respect of the first criterion, and that to a limited extent, since the values obtained are mean velocities over appreciable lengths of explosive. Only the photographic method can reveal the progress of the explosion; for example, it is clear that the records in figures 3 and 7 show unstable detonations in which the fluctuations in velocity would not have been disclosed in other methods of examination.

Instability of the detonation wave may arise directly from failure of the reaction in the wave front to sustain the shock front as a result, for example, of lateral losses in charges of small diameter or of undetected variations in the physical condition of the explosive. Cast T.N.T. is an explosive which is prone to unstable detonation due to variation in texture, even in charges of large diameter, and results which at first appear to be of a border-line nature may be ascribed to such variations. For the record shown in figure 12, plate 4, a cartridge of T.N.T , 2·2 cm. in diameter and 60 cm. long, was detonated by a tetryl primer of the same diameter and 10 cm. long. After a run of 50 cm. the high velocity and intense luminosity of the explosion suddenly ceased. The original negative (but not the reproduction) shows that flame continued beyond this point for a distance of about 5 cm. before fading out completely. After a moderately steady run, over the first 25 cm. (there was nearly a failure at the point marked by an arrow), the velocity was uniform and the wave evidently filled the cross-section of the cartridge; in this phase the wave could therefore be considered stable. It is suggested that at the 50 cm. point there was a variation in the physical texture of the explosive which was sufficient in the particular conditions of the experiment to cause a rapid fading of the detonation wave.

Persistence of uniformity of velocity is an important criterion, since unstable regimes of abnormal velocity may be induced temporarily by a strong primer; in figure 13, plate 4, a cartridge of Rounkol 12 cm. long is detonated by a pellet of 30:70-tetryl/T.N.T. (7025 m./sec.), and the velocity falls only gradually to the stable velocity of Rounkol (1850 m./sec.) over a distance of 8 cm.

There is some difference of opinion as to whether, for a specified explosive in a given diameter, only the highest uniform velocity attainable should be regarded as that of detonation, or whether a lower velocity can be so regarded provided it is uniform. We prefer the former view and regard the attainment of the highest value as a fourth criterion of detonation, and intermediate values as belonging to pre-detonation phases such as are obtained in certain explosive gas mixtures (Payman 1928). When the second view is held, explosion at a low uniform velocity is accepted as detonation when it persists for an appreciable distance, even though a higher velocity may also be attained later in the same cartridge. Well known amongst those which give two such regimes are the gelatinous explosives which contain a high proportion of nitroglycerine. The velocity of detonation of nitroglycerine is given as 7820 m./sec. and the low rate as 1590 m./sec. (Jones 1928). Figure 14, plate 4, shows the two regimes with N.S. gelignite, a blasting explosive containing 60 % of nitroglycerine. Here the high speed or, as we prefer to say, the velocity of detonation, is attained abruptly only towards the extreme end of the cartridge. A similar sequence is shown in figure 15, plate 4, obtained with crystalline tetryl (density 1·0) cartridged in cellophane and initiated by a no. 6 detonator; similar records are obtained with other explosives when a weak source of initiation is used. Both the cartridges of gelignite and tetryl detonate throughout at the full velocities if initiated by a sufficiently powerful primer.

ACCURACY OF THE PHOTOGRAPHIC METHOD

Since the Buxton rotating-mirror camera with its high writing speed of 360 m./sec. was first brought into use (Payman, Shepherd & Woodhead 1937), efforts have been made to improve its accuracy by the elimination of instrumental imperfections. Special emphasis is now laid on the use of (1) a single-sided mirror as thin as possible having regard to the stresses set up at high speeds of rotation; (2) a suitable timing device to ensure that the record falls on a selected portion of the film where there is negligible optical distortion; (3) a camera magnification (about 1/13 with the writing speed used) to give the flame trace a slope of about 45°, at which measuring errors are a minimum, and (4) a tuning-fork, the frequency of which can be regularly checked *in situ*, as the basis of the camera stroboscope. In addition, the method of assessing the records is important; ten separate measurements of each record are made by means of a vernier protractor, and the mean of the angles so determined is used to calculate the velocity. Agreement between results thus obtained by different observers is very close. Records in which the rates are not uniform are measured by an alternative method involving the use of a travelling microscope.

To assess the reproducibility of the velocity obtainable under controlled conditions, measurements were made with a crystalline T.N.T. of 98·5 % purity (set-point: 80·25° C). The charge was packed to a uniform density in cellophane tubes 3·17 cm. in diameter and 30 cm. long by supporting the wrapper in a metal tube and adding weighed quantities of the explosive. Each increment was compressed to the correct density by means of a marked wooden rammer. Confirmation of the suitability of the method was obtained by colouring alternate layers in trial cartridges and checking the length of each layer. The over-all cartridge densities varied between 0·99 and 1·01. The fineness of the explosive was such that 98 % passed through a no. 25 B.S.S. sieve and remained on no. 120, the sides of the mesh being 0·125 and 0·600 mm. respectively. The cartridges were primed with a 7 g. tetryl pellet and a no. 6 copper detonator. The results are given in the first column of table 1. From

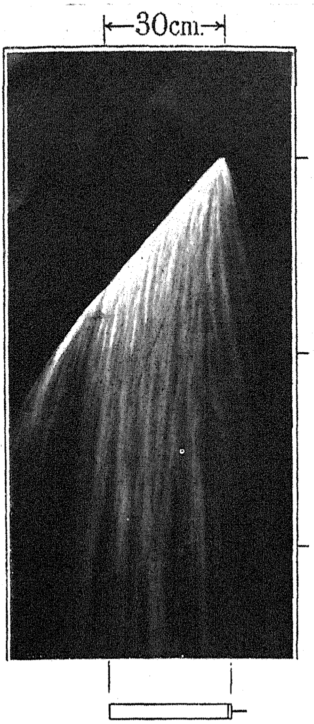


FIGURE 1. T.N.T. powder.

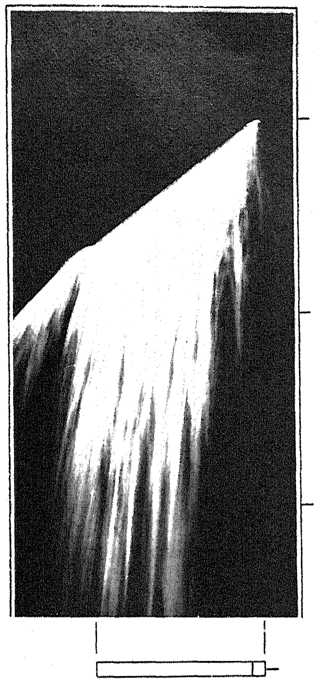


FIGURE 2. T.N.T. creamed (stable detonation).

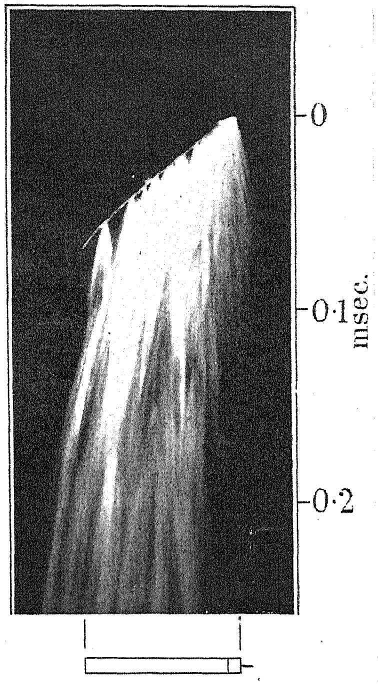


FIGURE 3. T.N.T. poured-clear (unstable detonation).

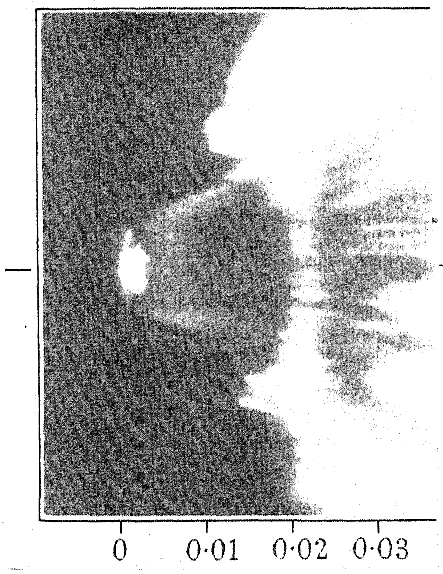


FIGURE 4. End-flame after unstable detonation.

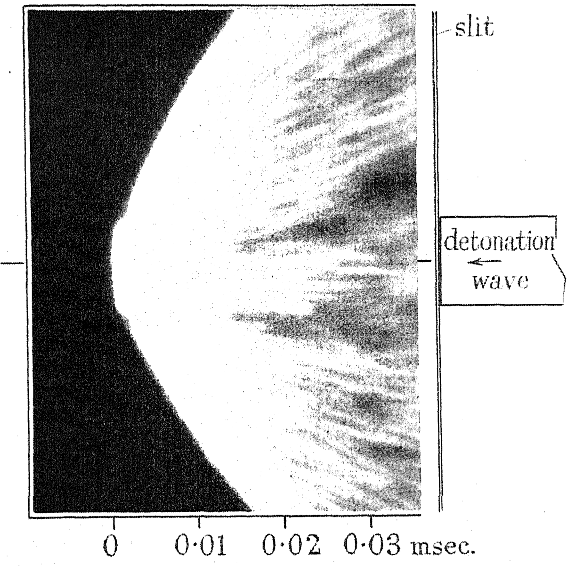


FIGURE 5. End-flame after stable detonation.

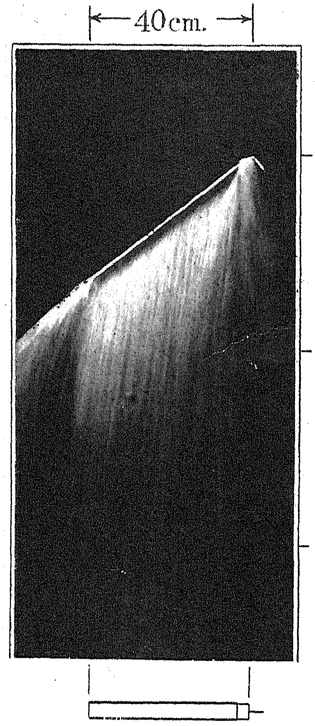


Figure 6. Creamed t.n.t. in a glass tube.

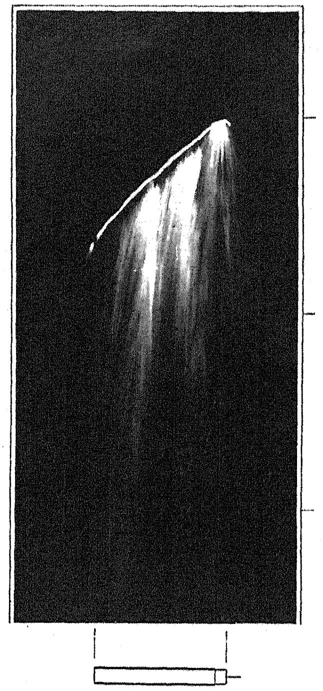


Figure 7. Poured-clear t.n.t. in a glass tube.

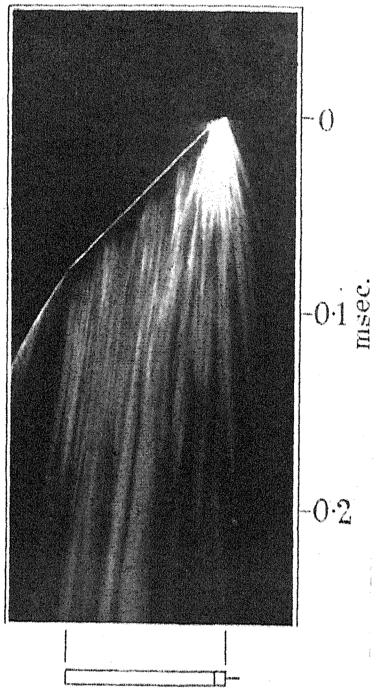
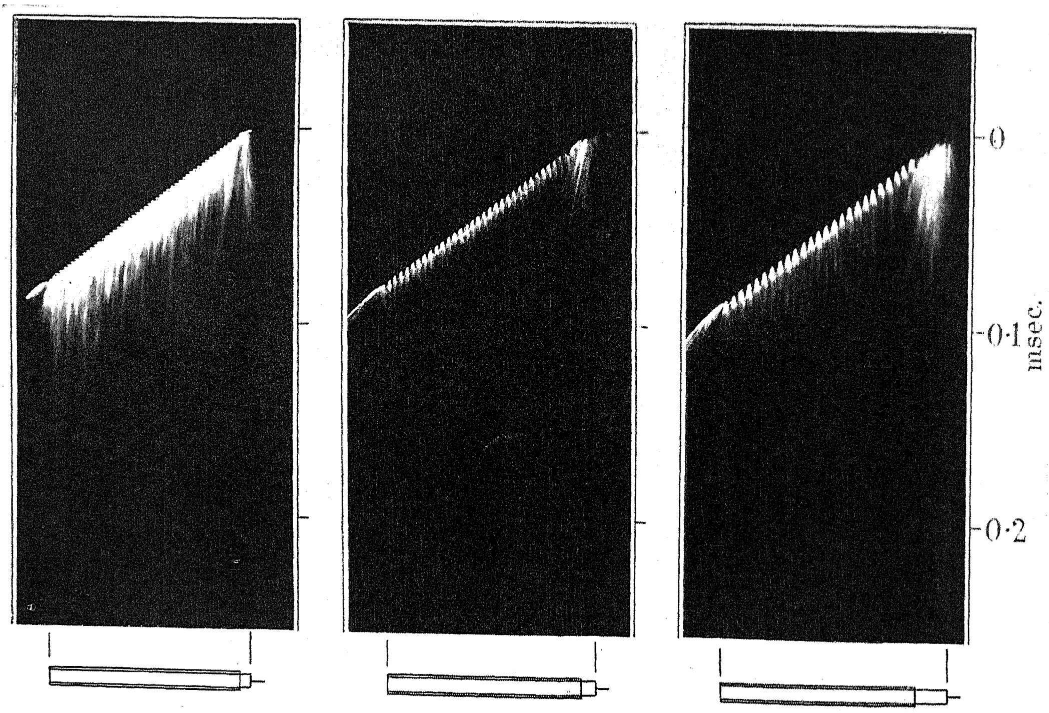


Figure 8. 50 : 50 amatol, unconfined.



Figures 9, 10, 11. Creamed t.n.t. in steel tubes with various perforations.

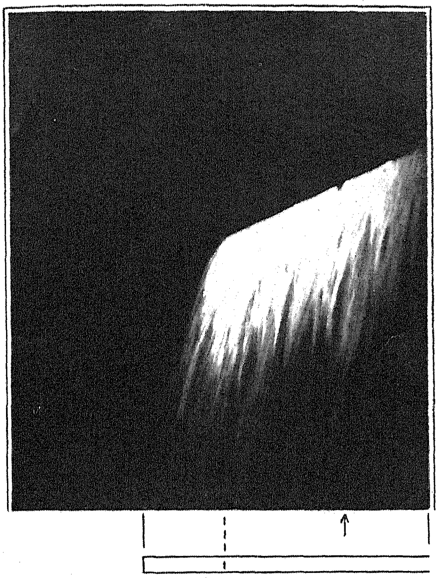


FIGURE 12. Fading of detonation in poured-cloudy T.N.T.

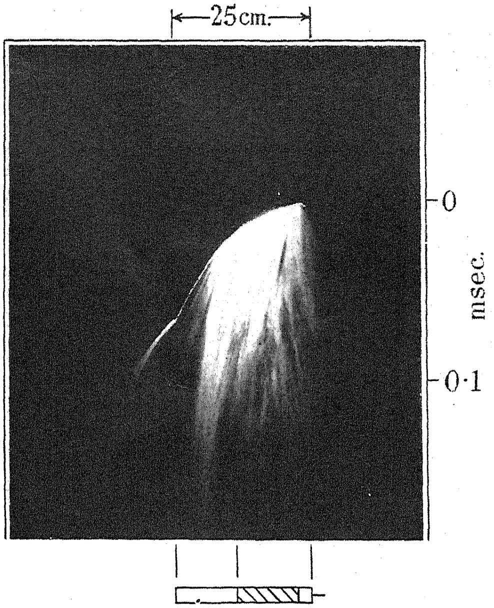


FIGURE 13. Composite cartridge.

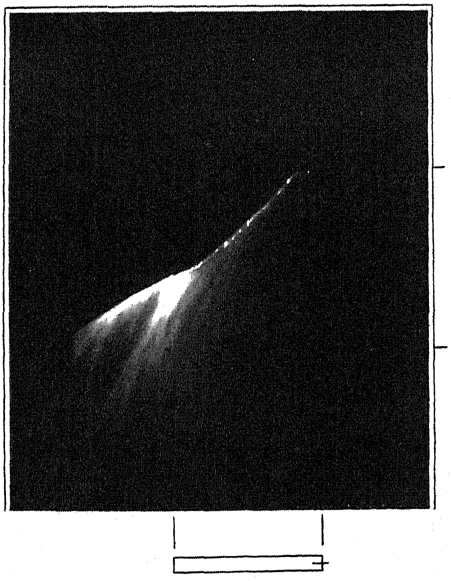


FIGURE 14. N.S. gelignite.

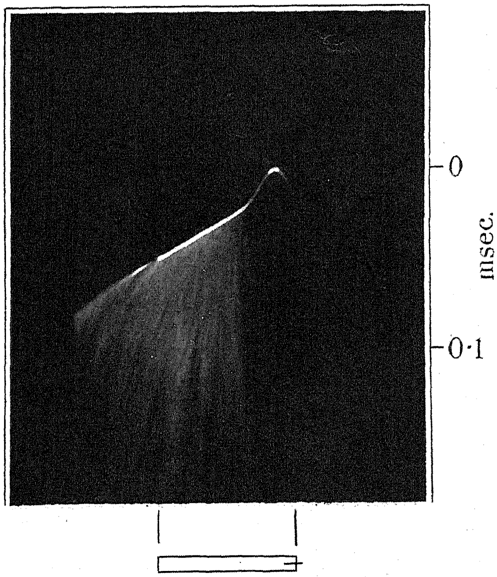


FIGURE 15. Tetryl powder.

TABLE 1. VARIOUS EXPLOSIVES: VELOCITY OF DETONATION OF CYLINDRICAL CHARGES, 3·17 CM. IN DIAMETER, UNCONFINED

explosive …	T.N.T.	picric acid	P.E.T.N./T.N.T. (50 : 50)	tetryl	tetryl/T.N.T. (5 : 95)		tetryl/T.N.T. (40 : 60)		
density (g./cm.^{3}) …	powder 1·00	cast 1·71	cast 1·677	pressed 1·506	cast 1·590		cast 1·62		
							(1)	(2)	(3)
number of shots fired	12	10	10	10	20		10	11	12
measured velocities (m./sec.)	4675	7296	7571	7175	6877	6848	7150	7133	7149
	4660	7275	7545	7170	6876	6843	7148	7131	7136
	4656	7275	7538	7162	6869	6842	7145	7127	7133
	4651	7274	7533	7155	6867	6837	7135	7120	7132
	4646	7273	7528	7153	6863	6835	7134	7117	7130
	4644	7263	7512	7151	6862	6832	7131	7110	7126
	4644	7257	7509	7150	6857	6832	7129	7108	7124
	4643	7239	7500	7148	6857	6829	7129	7096	7124
	4642	7235	7500	7142	6852	6828	7087	7094	7122
	4639	7227	7486	7135	6849	6826	7084	7092	7116
	4638							7080	7116
	4632								7115
mean velocity	4650	7260	7520	7150	6850		7127	7110	7127
range	43 *0·93*	69 *0·95*	85 *1·13*	40 *0·56*	51 *0·74*		66 *0·93*	53 *0·75*	34 *0·48*
standard deviation	12 *0·26*	22 *0·30*	26 *0·35*	12 *0·17*	16 *0·23*		23 *0·32*	18 *0·25*	10 *0·14*
standard error	3 *0·07*	7 *0·10*	8 *0·11*	4 *0·06*	4 *0·06*		7 *0·10*	5 *0·07*	3 *0·04*

twelve determinations of the velocity a mean value of 4650 m./sec. was obtained, the extreme values, 4632 and 4675 m./sec., giving a range of 43 m./sec., or 0·93 % of the mean value. The standard deviation of all the results was 12 m./sec. (0·26 %) and the standard error 3 m./sec. (0·07 %). It is evident from the photographic records (see figure 1) that stable detonation was attained.

Measurements with explosives giving high velocities of detonation were then made. Cylindrical pellets of five high-density explosives were prepared with care to ensure uniformity; for each experiment three pellets, each 3·17 cm. in diameter and 11·5 cm. long, were attached together in a column which was primed with a 28 g. pellet of pressed tetryl and a no. 6 electric detonator. Two of the explosives, picric acid and tetryl, were pure compounds, two were mixtures of tetryl and T.N.T. in the proportions 5 : 95 and 40 : 60 by weight respectively, and the fifth was a mixture of equal parts by weight of pentaerythritol tetranitrate (P.E.T.N.) and T.N.T. With the exception of the tetryl, the explosives were cast by us with a technique designed to ensure a uniform density. The tetryl was used in the form of pellets, 2·5 cm. long, specially pressed from pure granular material by the Explosives Division of I.C.I. Ltd.; the range of densities in a large sample of pellets was very small, namely, 1·5058 ± 0·0151.

Stable detonation was again attained with all the explosives, the records being of the type shown in figure 2, plate 2. The full results are given in table 1, the italic figures in this and subsequent tables being percentages of the mean velocities. Between every pair of explosives there is a significant difference in the mean velocities of detonation.

The narrowest range of velocities (about 0·5 %) occurs with pressed tetryl, the ranges for the other four explosives being of the order of 1 %. The density of each composite charge was measured and the variations from charge to charge were so small, ± 0·01, that they could not be positively associated with the variations in velocity observed with any one explosive.

The highest density of the picric acid used by Kast (see Stettbacher 1933, p. 61) was 1·69 and he obtained a velocity of detonation of 7285 m./sec., compared with our density of 1·71 and a mean velocity of 7260 m./sec.

The results of three series of determinations with cast 40 : 60-tetryl/T.N.T. made over a period of eight months are shown separately in table 1. The experimental conditions differed for the three series only in that the lengths of the composite cartridges were different (30, 23 and 35 cm., respectively), and in that the T.N.T. was taken from three batches of grade I quality; in no batch, however, was the percentage of α-T.N.T. less than 98·2. The purity of the tetryl was 99·8 %. Two observations may be made: the velocities of detonation in series 2 are on the whole lower than those in the two other series, and the scatter of the results is narrowest in series 3. These effects appear to be due to the use of long cartridges (35 cm.) in series 3 and comparatively short ones (23 cm.) in series 2. A statistical treatment of the results led to the conclusion that there is no significant difference between the mean velocities of detonation in series 1 and 3 at a probability of 0·05; there is a significant difference between the mean values obtained in series 2 and 3 but not in series 1 and 2. The greatest weight should therefore be attached to the results in series 3. These considerations suggest that for accurate reading of the records cartridges should not be less than 30 cm. long.

For the purposes of a further check of the reproducibility, experiments were made with granular P.E.T.N. which was available in a very high degree of purity. The fineness was such that 96·6 % passed through a no. 25 B.S.S. sieve and remained on no. 120, the natural packing density being 1·17; the same material was also pressed into short pellets at a mean density of 1·51. Cartridges 3·17 cm. in diameter and 35 cm. long were used, five determinations being made at each of the two densities. The two mean velocities were found to be 6110 and 7520 m./sec. respectively, the corresponding standard errors being 2 and 5 m./sec. The velocities agree fairly closely with those which Friederich (1933) determined by the Dautriche method, 6130 and 7460 m./sec. respectively.

We consider that these results are close to the absolute values for the velocities of detonation and to be more accurate than previously published figures obtained by the Dautriche method. It is, however, difficult to compare the relative reliabilities of the Dautriche and the present photographic methods, since the accuracy is usually expressed in terms of the range, but with little meaning when, as usually appears to have been the case with the Dautriche method, only few determinations were made. Results with the Dautriche method were stated by Kast & Günther (1919) to have a range of 3·8 %, and by Selle (1937) a range of 2 %. In their own more numerous trials Förg (1916) and Friederich (1931) each claim to have obtained a range of 1 %, but more recently Copp & Ubbelohde (1948), working with cast amatols, state that the method is capable only of providing values which lie within ± 1 % of the mean.

EFFECT OF PHYSICAL STATE OF T.N.T.

T.N.T. was usually obtained in the form of pale yellow flakes of 'grade I' purity; material of this grade has a specified set-point of not below 80·00° C. The set-point of α-T.N.T. is 80·90° C and the chief impurities responsible for a lower set-point are dinitrotoluene and isomers of T.N.T. and their approximate percentage is roughly equal to twice the difference between 80·90 and the set-point of the sample. A grade I material contains therefore 98 % or more of α-T.N.T.

The effect of crystal size

Castings of T.N.T. having different crystal sizes were produced by pouring the molten explosive at different temperatures and by arranging different rates of cooling. The following forms of casting were used in order of increasingly fine crystal structure:

(1) *Poured-clear* (coarse). Flaked T.N.T. is melted in a steam-heated vessel; at a chosen temperature (95° C), and whilst still clear, it is poured into long moulds and allowed to solidify.

(2) *Poured-cloudy* (medium). The molten T.N.T. is allowed to cool to the region of its set-point with continuous stirring until a slurry is formed which is then poured into moulds.

(3) *Creamed* (fine). The cooling melt is agitated with a mixer of the perforated disk type; when the slurry has formed, 10 % by weight of powdered T.N.T. (of such fineness that all passed through a no. 52 B.S.S. sieve) is added slowly; after further agitation the thickened melt is poured. Another form called *axially crystallized* consists of very long and massive crystals and is made by pouring clear into a mould arranged to cool slowly from one end. Samples of all these castings were examined microscopically by Drs F. J. Llewellyn and W. O. Williamson (Birmingham University), who are preparing a comprehensive description for publication.

The results given in table 1 emphasize the importance of using an ample length of charge. As it was found possible to cast the four types of T.N.T. uniformly in moulds 40 cm. or more in length, single charges of this minimum length were used in all later experiments. The axially crystallized castings could not be detonated unconfined even with a powerful primer. Poured-clear castings from some batches supplied could be detonated without confinement, from other batches they could not. Poured-cloudy and creamed castings detonated always, only a 7 g. tetryl primer being required for the creamed; a 28 g. tetryl primer usually sufficed for the poured-cloudy. This general experience of the initiation of detonation, confirmed by more quantitative work, showed that the sensitivity of T.N.T. becomes less with increasing crystal size of the cast.

TABLE 2. VELOCITY OF DETONATION OF CAST T.N.T. (1) POURED-CLEAR, (2) POURED-CLOUDY AND (3) CREAMED, IN CHARGES 3·17 CM. IN DIAMETER, 40 CM. LONG, UNCONFINED

	batch A		batch C	
T.N.T.	set-point: 80·32° C		set-point: 80·45° C	
casting method ...	poured-clear	creamed	poured-cloudy	creamed
density (g./cm.^{3}) ...	1·605	1·625	1·620	1·625
mean velocity (m./sec.)	6430	6850	6880	6880*
number of determinations	5	6	6	12
range	630	61	50	50
standard deviation	—	23	18	15
standard error	—	9	7	4

* See table 6 for individual values.

Due to the limited weights of the various batches of T.N.T. it was necessary to make comparisons of velocities in two series, between creamed and poured-clear from one batch, and between creamed and poured-cloudy from another batch. The charges, 3·17 cm. in diameter, were detonated unconfined. Table 2 shows that the poured-cloudy and the creamed castings give the same mean velocity in the diameter chosen for the experiments, but that the poured-clear gives values lower than those of the creamed and with a much greater range, as would be expected from the photographic records (cf. figures 2 and 3, plate 2).

The difference in velocities of detonation of creamed T.N.T. from the batches A and C is significant at a probability level of 0·1.

The effect of impurities

It has been important to find out how changing amounts of impurity in the explosive affect the velocity of detonation. Since creamed T.N.T. gave the highest velocity and the most stable detonation, it was used for a determination of the velocities of six batches of differing set-point, the one with the highest set-point (80·86° C) being specially purified by the Explosives Division of I.C.I. Ltd.

The velocities of detonation given in table 3 show a rough relationship with the set-points, the velocity being higher in the purer samples. The presence of the 1 % impurity is responsible for a reduction of about 100 m./sec. in the velocity of detonation; the differences between the various mean values are small, but they are significant (at a probability of 0·1) for all combinations except the first and last pairs. The value of the velocity of detonation in a pure explosive is of theoretical importance and the highest mean velocity observed (6950 m./sec.) can be accepted as the limiting value for pure α-T.N.T. at the stated density of 1·625.

TABLE 3. VARIOUS BATCHES OF T.N.T.: VELOCITIES OF DETONATION IN CREAMED CASTINGS, 3·17 CM. IN DIAMETER, 40 CM. LONG, DENSITY 1·625

batch ...	A	B	C	D	E	F
set-point (° C) ...	80·32	80·40	80·45	80·50	80·66	80·86
purity (% approx.) ...	98·8	99·0	99·1	99·2	99·5	99·8
mean velocity (m./sec.)	6850	6850	6880	6900	6940	6950
number of determinations	14	10	24	18	6	6
range	61	94	62	35	42	59
standard deviation	18	27	17	12	16	22
standard error	5	9	4	3	7	9

The effect of a much greater proportion of impurity is shown by experiments with a grade III T.N.T. (batch M, set-point 76·6° C) containing approximately 90 % α-T.N.T. The mean of six determinations of the velocity of detonation in creamed castings was 6680 m./sec., or 270 m./sec. below the highest mean value recorded in table 3, the highest density obtainable with this material being 1·57.

Effect of crystal size of ammonium nitrate in amatols

Results similar to those described above were obtained when crystals of ammonium nitrate in a range of finenesses were added to grade I T.N.T. to form amatols. The mixtures used were mainly in the proportions 50 : 50 and 60 : 40 of ammonium nitrate and T.N.T. respectively. In this type of explosive the crystals of ammonium nitrate retain their identity in the T.N.T. matrix. Velocities of detonation of unconfined cartridges, cast under carefully controlled conditions, have been determined as a part of a co-operative research with Dr J. I. O. Masson, F.R.S., of Sheffield University.* The ammonium nitrate used was selected from sieved fractions whose finenesses, defined by specific area, were determined by a transpiration method. Pellets, 3·17 cm. in diameter and 7 cm. long, were cast in lots of six in jacketed aluminium moulds; a stream of cold water was passed through the jackets to ensure rapid cooling and each casting was prodded to reduce sedimentation and cavitation. The ends of the solid pellets were faced on a lathe and a column of six was used for each determination of the velocity of detonation. Up to five separate determinations were made with each mixture. The microscopical examinations of the pellets will be described elsewhere and it suffices here to say that the influence of any stratification of the ingredients on the velocity of detonation was probably negligible, both because it was slight and because each charge consisted of a column of pellets. The influence of radial migration may not be wholly negligible since the rind is relatively rich in T.N.T. all the way along the cartridge.

---

* Dr Masson and Messrs Lloyd, Davies and Gwillim of his staff were responsible for the preparation of all the charges of amatol and for the determinations of the specific surface areas of the ammonium nitrate.

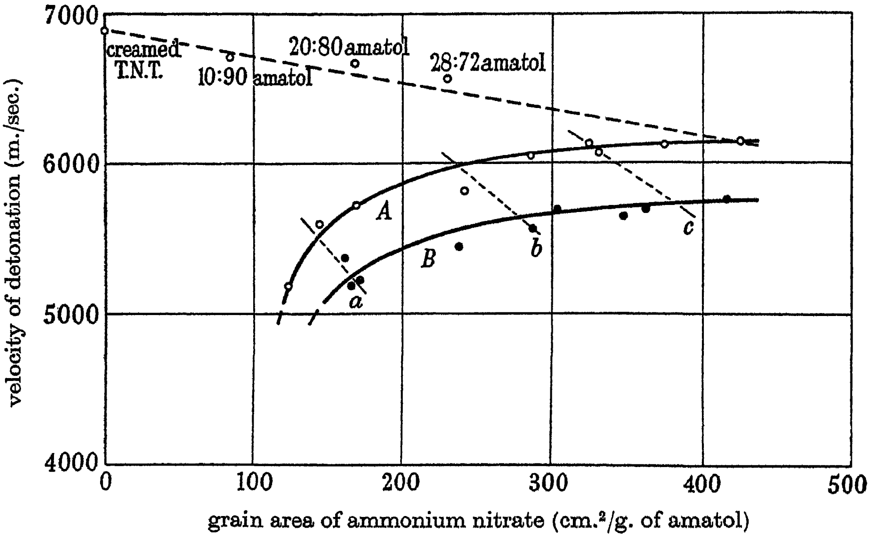


FIGURE 16. *A*, 50 : 50 amatols; *B*, 60 : 40 amatols. Unconfined cartridges 3·17 cm. in diameter. *a*, Service ammonium nitrate. *b*, R.O.F. milled ammonium nitrate. *c*, I.C.I. (Billingham) milled ammonium nitrate.

Table 4 contains all the determinations of the velocities of detonation. Amatols made with a given ammonium nitrate show a decrease in velocity with increase in nitrate content which is substantially linear, the value for creamed T.N.T. fitting well at the end of the curve. The latter is shown in the upper part of figure 16, where velocity is plotted against a parameter defining the 'grist', or fineness, of the ammonium nitrate in terms of the surface area in square centimetres per gram of amatol. Experiments with two amatols (50 : 50 and 60 : 40) show that a coarse grist, of small interfacial area, gives low and erratic velocities of detonation typical of unstable propagation (see figure 8), or failures to detonate at all. In gap-tests also, a cartridge of amatol needs to be closer to its fellow for detonation to be transmitted across the air-gap when the nitrate is coarse than when it is fine.

With progressively finer grain the velocity of detonation rises, approaching an upper limit beyond which greater fineness has no effect. On the graph in figure 16 the diagonal broken lines indicate that with factory made amatols the nitrates were not fine enough to cause the maximum velocity of detonation in unconfined charges.

TABLE 4. AMMONIUM NITRATES OF VARIOUS FINENESSES: VELOCITIES OF DETONATION IN CAST AMATOLS, 3·17 CM. IN DIAMETER AND 42 CM. LONG, UNCONFINED

amatol ...	0:100	10:90	20:80	28:72	50:50
grist* ...	—	84	168	230	420
velocity of detonation ...	6870	6700	6660	6560	6150

amatol 50:50 (density: 1·58 ± 0·03)		amatol 60:40 (density: 1·55 ± 0·03)	
grist	velocity	grist	velocity
420	6150	416	5770†
375	6120	363	5700
331	6070	348	5650
325	6140	304	5700
286	6050	286	5560‡
241	5820	238	5450
169	5730	172	5240
145	5600	166	5190
124	5190	162	5380
93 79	no detonation	93 90	no detonation

* Grist of this nitrate: 840 cm.^{2}/g. (through no. 200 B.S.S. sieve).

† Mean of two similar batches in which the 60 parts of nitrate consisted of 45 parts grist 840 cm.^{2}/g. and 15 parts grist 250 cm.^{2}/g.

‡ This is the mean of five determinations: 5585, 5582, 5554, 5550 and 5544 m./sec.; range 41 and standard deviation 18.

Experiments were made to test the suggestion that the velocity of detonation depends on the *average* grain area of the ammonium nitrate present rather than on the actual sizes of the individual crystals. Two batches of 60:40-amatol were made up with (i) a nitrate as supplied for Service use, and (ii) a nitrate consisting of fine and coarse materials in proportion to give the same surface area as that of the Service nitrate; the velocities are almost the same, namely 5240 and 5190 m./sec., respectively, showing that it is the average grain area on which the velocity mainly depends.

For an examination of the effect of ageing on the velocity of detonation, a 60:40 amatol was made from a Service nitrate sieved through a no. 25 B.S.S. sieve to remove aggregates (surface area 162 cm.^{2}/g. of amatol). Several hundred pellets were prepared in a short time under identical conditions so as to ensure a high degree of uniformity. The velocity of detonation was determined a few days after the preparation and again after storage for periods of 3, 6 and 12 months under dry temperate conditions, the velocities being 5380, 5270, 5280 and 5240 m./sec. respectively (each the mean of five determinations). The velocity dropped by about 100 m./sec. after the first 3 months and only by 140 m./sec. in the year.

H. Jones (1942) has suggested that, in the detonation of an explosive consisting of two or more compounds having different rates of reaction, there is erosion of the particles of the less reactive material by the hot gases released in the exothermic decomposition of the most reactive ingredient. From this hypothesis it follows that the duration of the reaction should be inversely proportional to the surface area presented by the more slowly reacting constituents. In amatols, for example, the ammonium nitrate reacts more slowly than the T.N.T., the over-all time of reaction is the time required for the complete erosion of the ammonium nitrate and is therefore inversely proportional to the total initial surface area of that component. These theoretical considerations imply an increase in velocity of detonation with increasing fineness, and in this respect are confirmed by our experiments. For, so far as the pure explosive is concerned, we have already seen that finely crystalline T.N.T. propagates detonation more easily, and in general has a higher velocity than when it is made up of large crystallites. Further, when ammonium nitrate is added to form the amatols, the velocity of detonation increases with increasing surface area of the nitrate. By plotting the results given in table 4 as the reciprocals of the velocity against those of the specific surface area, we find that, by extrapolating to infinite fineness, the maximum velocities of detonation for 50:50 and 60:40 amatols are respectively 6210 and 5860 m./sec.

EFFECT OF DIAMETER OF CHARGE

That the velocity of detonation of a given explosive is in general higher in wide cylindrical charges than in narrow ones seems to have been observed first by Berthelot (1885). More recent examinations of the effect have been concerned chiefly with explosives at comparatively low loading densities; for example, Parisot & Laffitte (1938), working with picric acid at a density of 0·9, found that the velocity of detonation was 2230 m./sec. when the diameter was 0·54 cm. and 4575 m./sec. when it was 1·7 cm. Since much smaller changes in velocity with diameter were expected and since the continuity of detonation was also in question, the photographic method of measurement was necessary in our work with high-density explosives such as cast T.N.T.

Preliminary experiments were made with two explosives in which detonation was readily established. Pressed tetryl pellets of a high degree of purity were available in diameters of 2·22 cm. (⅞ in.) and 3·17 cm. (1¼ in.). The mean velocity in unconfined charges of the smaller diameter was 6880 m./sec. (density 1·44), compared with 7150 m./sec. in the larger (density 1·50; see table 1). Trials were then made with a laboratory-cast mixture of tetryl and T.N.T. (30:70), density 1·613; mean velocities for the two diameters 0·9 and 3·17 cm., detonated unconfined, were 7020 and 7030 m./sec. respectively, the difference not being significant. In view of the importance of the effects of crystal texture on the velocity and stability of detonation in cast T.N.T. it was realized that the experiments should be conducted with castings produced under strictly controlled conditions. It is the rate of cooling which controls the crystal size and the latter, even with a given method of casting, must therefore depend to some extent upon the diameter of the charge. Cylindrical charges of either poured-clear or poured-cloudy T.N.T. cast in moulds of different diameters show appreciably different grain sizes and the observed differences in velocity observed with the latter type are in part due to this unavoidable complication. This difficulty is not so serious with creamed T.N.T., the crystal size of which is not seriously affected by variations in the diameter of the casting; with all castings, however, there is a 'chilled margin' of very small crystals produced during the first rapid solidification of the explosive in contact with the wall of the mould. The thickness of the margin is of the order of 0·25 mm. and any effect it may exert would be the greater the smaller the cylinder. Density also seems to change with diameter and, with creamed T.N.T., there is a small but steady decrease as the diameter is decreased; this appears to be unavoidable with the methods of casting used, and the use of warmed moulds was not found to lead to any greater over-all uniformity. The T.N.T. used was a grade I flake of set-point 80·50° C (batch D); the castings were 50 cm. long, except the largest (3·17 cm. diameter) which were 40 cm. long. They were produced by pouring into brass tubing 0·15 cm. thick. Since it had previously been found that excessive power in the priming charge had no disturbing effect on the steady velocity in a charge capable of detonation, primers of ample strength were always used, the diameter of the primer being the same as that of the charge.

The poured-clear castings of T.N.T. from batch D did not detonate even in the largest diameter (3·17 cm.); similar castings from an earlier batch A with a rather lower set-point (80·32° C) had been successfully detonated although with a very wide range (see table 2). The poured-cloudy castings detonated in diameters of 3·17 and 2·54 cm., but not in diameters of 2·20 cm. or less. As shown in table 5, the range and the standard deviation of the individual results are greater with a diameter

TABLE 5. DIFFERENT DIAMETERS OF CAST T.N.T.: POURED-CLEAR, POURED-CLOUDY AND CREAMED, SET-POINT 80·50° C. (BATCH D); UNCONFINED, LENGTH 50 CM.

charge diameter (cm.) ...	1·26	1·66	1·90	2·20	2·54	3·17*
T.N.T., poured-clear						no detonation
T.N.T., poured-cloudy:						
density (g./cm.^{3})				no detonation	1·615	1·620
measured velocities (m./sec.)					6698	6919
					6681	6915
					6681	6910
					6637	6905
					6562	6898
					6560	6890
mean velocity					6640	6910
range					138	29
standard deviation					62	11
standard error					25	4
T.N.T., creamed:						
density (g./cm.^{3})	no detonation	1·605	1·610	1·615	1·615	1·620
measured velocities (m./sec.)		6522	6754	6854	6904	6919
		6507	6748	6846	6891	6911
		6482	6742	6814	6884	6909
		6455	6741	6812	6883	6894
		6454	6738	6810	6876	6888
		6406	6719	6801	6874	6884
mean velocity		6470	6740	6820	6890	6900
range		116	35	53	30	35
standard deviation		42	12	22	11	14
standard error		17	5	9	4	6

* 40 cm. long.

of 2·54 cm. than of 3·17 cm., indicating that with the former the detonation is tending to be more sensitive to undetected variations in texture which may cause fading; this conclusion is supported by the more marked saw-tooth appearance of the photographic records obtained with the 2·54 cm. charges.

With creamed T.N.T., detonation is obtained with diameters down to 1·66 cm., the range here being greater than with diameters of 1·90 cm. and upwards. The velocities in the two larger diameters, 2·54 and 3·17 cm., are approximately the same. Statistical analysis of all the results in table 5 shows that the mean velocities of detonation are significantly different with the exceptions of that of creamed T.N.T. in the diameter 3·17 cm. in relation to those of creamed T.N.T., 2·54 cm., and of poured-cloudy T.N.T., 3·17 cm.; only within these groups do the experimental variations overshadow the variations due to the change either in the explosive or in the diameter of the charge.

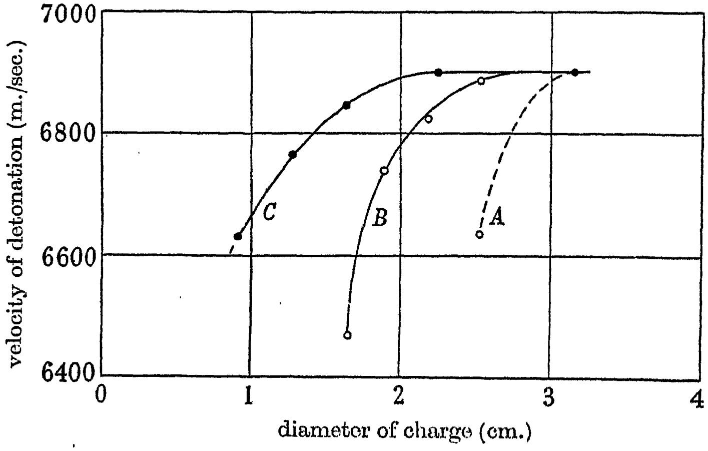


FIGURE 17. The velocity of detonation in cylinders of T.N.T. of various diameters. *A*, poured-cloudy, unconfined; *B*, creamed, unconfined; *C*, creamed, confined in steel tubes.

These determinations show that increase in diameter up to a certain limit, which we have called the *limiting diameter*, results in an increase up to the (limiting) velocity of detonation; below a certain diameter, termed the *boundary diameter*, there is no detonation. The values of these parameters vary with the nature and the physical state of the explosive and they are different for confined and unconfined charges. The boundary diameter provides a measure of the sensitivity to propagation of detonation. The mean results are shown graphically in figure 17; the curve for poured-cloudy T.N.T. *A* is assumed to be similar in shape to that for creamed T.N.T. *B*. Both castings have the same limiting velocity of detonation, but differ in their limiting and boundary diameters and in their velocities of detonation in cartridges below the limiting diameter. It is evident that differences in detonation properties, either non-existent or not measurable in wide cartridges, are marked in narrow ones.

Any effect of the chilled margin on the velocity of detonation might be expected to be the more pronounced in the smaller diameters. A few determinations were therefore made with narrow cylinders of creamed T.N.T. from which the margin had been removed by glass paper, and others cast to the same final diameter and therefore still possessing a margin. There was no evidence that either the velocity of detonation or the boundary diameter of the creamed castings was affected by the presence of a margin.

With crystalline T.N.T. at a loading density of 1·10, the limiting velocity was not reached in unconfined cartridges 3·17 cm. in diameter, although later experiments with confined cartridges indicated that the limit (about 5250 m./sec.) was not far above that obtained in this diameter. The boundary diameter was 1·66 cm. (no detonations having been obtained at 1·45 cm.) with a velocity of detonation of 4350 m./sec.; at diameters of 1·90, 2·20 and 3·17 cm. the velocities were 4620, 4740 and 5000 m./sec. respectively, each being the mean of three determinations. The crystalline T.N.T. used for these experiments, although of grade I quality, was less pure than that of previous batches, having a set-point of 80·05° C. Its fineness, such that 97·5 % passed through a no. 25 B.S.S. sieve and remained on no. 120, was very similar to that of the crystalline T.N.T. (set-point 80·25° C) used for the velocity determinations recorded in the first column of table 1. The lower value of the velocity (at 3·17 cm. diameter) in the first set of experiments can be ascribed mainly to the lower loading density used, that is, 1·00 compared with 1·10. The most curious feature of the results at different diameters is that the boundary diameter (1·66 cm.) is the same as that with the higher density cast explosive (table 5—creamed T.N.T.). In view of the lower purity of the crystals compared with that of the flake used for the experiments recorded in table 5 (set-point 80·50° C), the crystalline material was used to prepare six creamed charges in each of the diameters 1·45 and 1·66 cm. All the charges of the latter diameter detonated giving a mean velocity of 6770 m./sec. Three of those 1·45 cm. in diameter failed to detonate; the mean velocity of the remainder was 6560 m./sec. Thus the T.N.T. of the lower set-point had a boundary diameter quite close to that of the batch with the higher set-point, and the similarity of boundary diameters at the very different loading densities of 1·1 and 1·6 can be accepted. It is known that this explosive is considerably more sensitive to initiation when loaded at low than at high density and it would therefore appear necessary to distinguish between initiation sensitivity and propagation sensitivity.

EFFECT OF CONFINEMENT

The effect of confinement is to reduce loss of energy in the wave front and is similar to that of increasing the diameter of the charge. Thus, the velocity of detonation is raised by confinement only at diameters below the limiting diameter. With perfect confinement, by a rigid non-conducting material, the velocity of detonation should be independent of diameter, but even a thick steel casing does not provide this idealized condition. For we have found that there is a limiting diameter (smaller than that for bare charges) below which the slight losses to the confining material reduce the velocity of detonation below the limiting velocity. To achieve this limiting velocity, the lateral loss must not be felt across the whole cross-section of the charge and a central core must detonate under the perfect confinement provided by the next outer layer also undergoing detonation. That is to say, at least the central portion of the detonation wave must be plane-fronted.

The experiments with confined charges were made with creamed T.N.T. in a charge diameter of 3·17 cm. and with confinement provided by cold-drawn mild steel tubes of 0·32 cm. wall thickness. The results given in table 6 were obtained first of all to test the accuracy of the drilled tube method of measurement and they compare favourably with those for a similar series of unconfined charges of T.N.T. in the same diameter, which, it should be noted, is above the limiting diameter. The over-all range of the individual results is again less than 1 % of the mean, and the standard error is below 0·1 %.

It is convenient to express the confinement, S, in terms of the mass of the confining material in grams per square centimetre of lateral surface area of the cylindrical charge; the value of S in the experiments in table 6 is 2·75 g./sq.cm.

TABLE 6. BARE AND CONFINED CHARGES OF CREAMED T.N.T.: BATCH C, SET-POINT 80·45° C., 3·17 CM. IN DIAMETER, 40 CM. LONG, DENSITY 1·625

	bare		confined in steel tube, 0·32 cm. wall thickness	
number of shots fired	12		12	
measured velocities (m./sec.)	6918	6875	6916	6887
	6897	6870	6902	6883
	6893	6869	6902	6879
	6885	6869	6901	6861
	6884	6869	6900	6860
	6877	6868	6895	6856
mean velocity	6880		6890	
range	50	*0·73*	60	*0·87*
standard deviation	15	*0·22*	19	*0·27*
standard error	4	*0·06*	5	*0·08*

Effect of material of confining tube

Experiments were made with T.N.T. (batch A, set-point 80·32° C) cast in tubes of steel, lead and glass, each 3·17 cm. in internal diameter and of 0·32 cm. wall thickness. The value of S of the three materials was 2·75, 3·62 and 0·33 g./sq.cm., respectively. Since the maximum velocity of detonation of creamed T.N.T. is attained in unconfined cylinders below this diameter, it was not to be expected that confinement would influence the velocity, and, in fact, the mean speeds were all found to lie between 6845 and 6855 m./sec.

We have seen that the detonation in poured-clear T.N.T. is uncertain and, when obtained, the velocity in unconfined charges 3·17 cm. in diameter is lower than that in similar charges of creamed T.N.T. (see table 2); the establishment of detonation in the poured-clear casting becomes easier with confinement, and experiment shows that the velocity (batch B) rises from 6430 to 6700 m./sec. when confined in either the glass or the lead tubes and 6810 m./sec. in the steel (each the mean of six determinations). Hence, confinement in lead has less influence than confinement in an equal thickness of steel, despite the lower mass of the latter. The value of 6810 m./sec. obtained with poured-clear T.N.T. in the steel tube is below that with creamed T.N.T. from the same batch (6850 m./sec.), but the difference is probably due to a difference in the respective densities (1·605 and 1·625 respectively). Although the value of S

with glass is less than one-tenth that with lead, the same velocities of detonation are obtained; however, the pre-detonation period is much longer in the glass tube and the rate of propagation is low and unsteady over the first 10 to 15 cm. of the charge.

Berthelot (1885, p. 565), working with cartridges of gun-cotton 4 to 6 mm. in diameter, found that confinement in pewter tubes gave a higher velocity of detonation (5916 m./sec.) than confinement in tubes of lead (5200 m./sec.); he suggested that this difference occurs 'peut-être parce que le premier métal résiste un peu plus longtemps que le second à l'effort de l'explosion qui détruit le tube'. H. Jones (1947), in a theoretical treatment, states that it is not a serious approximation to neglect the strength of the confining material, since the yield stress, even of steel, is small compared with stresses arising near the front of the detonation wave, and that the compressibility of the material is an important factor. For thin (laminar) casings, which provide only partial confinement, it is the mass per unit area which decides the increase in the velocity of detonation; for thick casings, compressibility of the confining material is the physical property involved. Copp & Ubbelohde (1948) have developed this view and state that when the thickness of the wall is such that the shock wave in the casing material does not reach the outside wall within the reaction time, compressibility is the controlling factor in deciding the efficacy of a given material.

It is interesting to look at the results for steel, glass and lead in the light of this theory. The compressibility of lead and glass is similar and several times greater than that of steel. The wall thickness (0·32 cm.) is not dissimilar to the length of the reaction zone, and, as the velocity of sound in glass is not much below the velocity of detonation in the explosive, it would seem that the confinement may be classed as thick. Hence, the effect of compressibility rather than that of mass should preponderate, and this is the result observed.

Effect of diameter with confined charges

In these experiments the charge was cast directly into each individual piece of steel tubing. We were unable to obtain tubes of various internal diameters and of such thickness that the value of S was maintained constant. In most of the determinations, however, S varied only between 1·6 and 2·2 g./sq.cm. Cold-drawn mild steel tubing was used for the 3·17 cm. charges; the narrower tubing was butt-welded and had a rather rougher inner surface. The results of the individual determinations with poured-clear and creamed castings of T.N.T. (batch D, set-point 80·50° C) are given in table 7.

Poured-clear castings could be detonated in all the steel tubes, but the range and the standard deviation of the results with the narrowest tube (diameter: 0·94 cm.) suggested that this was near the boundary diameter. The limiting diameter seemed to be reached at 1·65 cm., but the maximum velocity was only 6830 m./sec. The densities fluctuated between 1·590 and 1·610, varying inversely with changes in crystal size through the range of diameters. For example, owing to the more rapid cooling, the crystals were less coarse in the 0·99 cm. tube than in the 1·29 cm. tube, and the former gave a rather higher density and a corresponding higher velocity of detonation; when a narrow casting (0·94 cm. diameter) was made with the tube

TABLE 7. EFFECT OF CONFINEMENT IN STEEL TUBES OF VARIOUS DIAMETERS: POURED-CLEAR AND CREAMED T.N.T. BATCH D, SET-POINT 80·50° C, LENGTH 50 CM.

steel tube:							
internal diameter (cm.) ...	0·94*	0·99	1·29	1·65		2·26	3·17†
wall thickness (cm.) ...	0·23	0·18	0·20	0·23	0·08	0·20	0·32
S (g./sq.cm.) ...	2·17	1·61	1·86	2·07	0·66	1·73	2·75
T.N.T., poured-clear:							
density (g./cm.^{3})	1·590	1·610	1·605	1·605		1·590	1·600
measured velocities (m./sec.)	6383	6794	6783	6855		6867	6839
	6343	6794	6774	6835		6846	6838
	6335	6786	6769	6826		6837	6825
	6288	6781	6764	6821		6817	6823
	6225	6774	6745	6817		6815	6816
	6208	6771	6742	6800		6791	6809
		6770	6736				
		6768	6727				
		6765	6721				
		6725	6700				
mean velocity	6300	6770	6750	6830		6830	6830
range	175	69	83	55		76	30
standard deviation	69	20	26	18		27	12
standard error	28	6	8	7		11	5
T.N.T., creamed:							
density (g./cm.^{3})	1·600		1·605	1·605	1·605	1·610	
measured velocities (m./sec.)	6659		6788	6871	6804	6915	
	6641		6775	6866	6790	6913	
	6634		6773	6856	6787	6913	
	6630		6772	6847	6782	6902	
	6620		6760	6822	6779	6890	
	6591		6735	6820	6755	6887	
mean velocity	6630		6770	6850	6780	6900	
range	68		53	51	49	28	
standard deviation	23		18	22	16	12	
standard error	9		7	9	7	6	

* Tubes preheated to 75° C for poured-clear and to 50° C for creamed T.N.T. † 40 cm. long.

wall preheated to 75° C the velocity was considerably lower owing to the coarse crystallization resulting from the slower cooling. When the mould was at room temperature the density of the narrowest cylinder was the highest in the whole series; when it was warmed to 75° C before the explosive was poured the density was the lowest observed.

The results with creamed T.N.T. were more regular. In order to prevent premature setting of the creamed melt, it was necessary to warm the narrowest tube to 50° C before pouring; this had no detectable effect on the crystal size. There was a small but steady decrease in density as the diameter was decreased. The velocity of detonation in the 2·26 cm. tube was the same as the maximum velocity in unconfined charges 3·17 cm. in diameter (see table 5) and the limiting diameter for confined charges is therefore not more than 2·26 cm. The range and the standard deviation of the velocities in the 0·94 cm. tube are a little greater than those in the wider tubes, but not sufficiently to suggest that the boundary diameter was near. Two series of experiments were carried out with charges 1·65 cm. in diameter confined in steel tubing of two different wall thicknesses, the values of S being 2·07 and 0·66 g./sq.cm. respectively. With the weaker confinement there was a reduction of about 60 m./sec. in the mean velocity of detonation. The effect of diameter with confined and unconfined charges of creamed T.N.T. of batch D is shown graphically in figure 17. The curve B for the unconfined explosive is complete, no detonations having been obtained in diameters less than 1·66 cm., whereas the curve for the confined charges C is incomplete in that the boundary diameter may well be below 0·94 cm., six detonations having been obtained at that point. The curves clearly reach a common maximum of velocity at a diameter of about 2·5 cm. where neither confinement nor further increase in diameter affects the velocity of detonation.

The significance of the boundary diameter lies in the fact that the magnitude of the energy loss through lateral expansion during the reaction period is so great that there in only just sufficient residual energy in the direction of propagation for the maintenance of a detonation wave. On further small reduction in diameter the rarefaction waves set up by the lateral expansion of the product gases are able to overtake the wave front, causing here a loss of amplitude and therefore of the initiating power of the detonation pulse. As the wave progressively degrades, detonation dies out; this sequence of events is observed with both confined and bare charges, though, as would be expected, at a much smaller diameter for the former than for the latter. Evaluation of these critical conditions gives a measure of sensitivity to propagation in a specified explosive, and it is clear that, if the lateral loss can be assessed, then the critical energy necessary for the maintenance of detonation can also be found.

It has been shown that the detonation properties of T.N.T. differ from batch to batch, and hence it is not possible to lay down standard figures for the constants of detonation. In table 8 are summarized the results for batch D, which has a set-point of 80·50° C and is therefore of high quality. An interesting point emerges with regard to the effect of crystal size, for it is clear that under adequate conditions of confinement and diameter the velocity of detonation of the three above varieties of casting tends towards the limiting value of 6950 m./sec. obtained with the purest T.N.T. available (cf. table 3).

TABLE 8. LIMIT DATA OF DIAMETERS AND VELOCITIES OF DETONATION OF CAST T.N.T., BATCH D, SET-POINT 80·50° C

		poured-clear		poured-cloudy		creamed	
	casting	bare	confined	bare	confined	bare	confined
boundary:	diameter (cm.)	—	$\ngtr 0{\cdot}9$	2·5	—	1·7	$\ngtr 0{\cdot}9$
	velocity (m./sec.)	—	6300	6400	—	6470	6600
limiting:	diameter (cm.)	—	1·7	$\sim 3{\cdot}0$	—	2·7	2·2
	velocity (m./sec.)	—	6830	6900	6900	6900	6900

The results with the poured-clear castings show that although the limiting velocity is slightly below that of the creamed variety, the limiting diameter (confined) of this apparently more insensitive material is smaller.

Confined charges of axially crystallized T.N.T.

Experiment showed that it was possible to detonate this variety in a steel tube 3·17 cm. in diameter, the castings being made with T.N.T. from a batch L, having a set-point of 80·05° C. The velocity of detonation was 6450 m./sec. (range 180 m./sec., density 1·600), this being the mean value of twelve determinations. Five of the experiments were made with tubes of wall thickness 0·32 cm. and the remaining seven with tubes 0·64 cm. thick, S being 2·75 and 5·88 g./sq.cm., respectively. The over-all mean of the twelve determinations is acceptable because there was not a significant difference between the two sets of values, a fact which, in addition, indicates that the limiting velocity for this variety of T.N.T. had been reached. As the limiting velocity of detonation of this batch of T.N.T. when poured-clear was found to be 6870 m./sec., we have now to explain the fall of 420 m./sec. with the axially crystallized material. The poured-clear and axially crystallized varieties differ only in the direction of the crystal growth, both being poured at the same temperature. With the former the growth is radial and with the latter axial, and so the results suggest that there are characteristic values for the velocity of detonation according to the orientation of the crystals with respect to the direction of propagation. The dimensions of the crystals in each of these slowly crystallized castings are large in comparison with the thickness of the reaction zone (probably about 3 mm.). If, as has been suggested, there are differences in the rate of reaction along different crystallographic axes, corresponding differences in the velocity of detonation would be expected. As an extension of this hypothesis it may be suggested that propagation of detonation is stable along one crystal axis and unstable along another axis.

This work was carried out during 1943–46 in the Buxton laboratories of the Safety in Mines Research and Testing Branch of the Ministry of Fuel and Power, under an extra-mural contract with the Ministry of Supply; permission for publication has been granted by the Chief Scientist of the latter Ministry.

REFERENCES

Abel, F. A. 1873–4 *Proc. Roy. Soc.* **22**, 160–171.
Berthelot, P. E. M. 1885 *Ann. Chim. Phys.* **6**, 556–574.
Copp, J. L. & Ubbelohde, A. R. 1948 *Trans. Faraday Soc.* **44**, 646–658.
Dautriche, H. 1906 *C.R. Acad. Sci., Paris*, **143**, 641–644.
Förg, R. 1916 *Z. Schiess u. Sprengstoffwesen*, **11**, 17–32 and 37–41.
Friederich, W. 1931 *Z. Schiess u. Sprengstoffwesen*, **26**, 221–223.
Friederich, W. 1933 *Z. Schiess u. Sprengstoffwesen*, **28**, 2–6.
Jones, E. 1928 *Proc. Roy. Soc.* A, **120**, 603–620.
Jones, H. 1942 Unpublished results in a Ministry of Supply Report.
Jones, H. 1947 *Proc. Roy. Soc.* A, **189**, 415–426.
Kast, H. & Günther, P. 1919 *Z. Schiess u. Sprengstoffwesen*, **14**, 120–124.
Parisot, A. & Laffitte, P. 1938 *Congr. Chim. Ind., C.R.* 18*me Congr.*, Nancy, pp. 930–936.
Payman, W. 1928 *Proc. Roy. Soc.* A, **120**, 90–109.
Payman, W., Shepherd, W. C. F. & Woodhead, D. W. 1937 *Pap. Safety Min. Res. Bd, Lond.*, no. 99.
Selle, H. 1937 *Z. Schiess u. Sprengstoffwesen*, **32**, 179–183.
Stettbacher, A. 1933 *Die Schiess und Sprengstoffe*, 2nd ed. Leipzig: J. A. Barth.

On the self-energies and cross-sections of orthodox quantum mechanics

By H. S. Green, *Edinburgh University*

(*Communicated by M. Born, F.R.S.—Received* 20 *August* 1948—
Revised 19 *October* 1948)

In this paper are introduced several novel techniques having as their objects, first, the gain of a new understanding of the divergence difficulties of orthodox quantum theory, and secondly, the systematic development of the quantum mechanics of fields in terms of the density matrix.

There is included a new presentation of a form of perturbation theory which has the outstanding advantages over the usual one that it is much more quickly convergent, and leaves the density matrix normalized at every stage so that the diagonal elements representing occupational probabilities cannot diverge. Exact general formulae are derived for self-energies and cross-sections for the purpose of examining their convergence properties.

The general theory of fields is developed *ab initio*, and it is shown that the use of the density matrix in place of the wave vector illuminates and simplifies the customary theory. The second quantization of the density matrix which follows throws unexpected light on the existence of particles with negative expectation values.

Finally, the whole theory is applied to the crucial example of the electron or positron in an electromagnetic field. It is confirmed that the self-energy of the electron in the orthodox theory cannot be made finite without the introduction of negative energy photons.

1. INTRODUCTION

The principal obstacle to progress in several branches of quantum mechanics and especially in quantum electrodynamics has long since been recognized to be the appearance of infinities quite early in the calculation of self-energies and cross-sections for scattering by the usual perturbation approximation methods. It is still unsettled whether the difficulties associated with these infinities arise from the limitations of the mathematical technique employed, or whether they are due to the inadequacy of the physical ideas and assumptions expressed by the orthodox theory. Assuming that the latter alternative is correct, several authors (e.g. Born & Rumer 1931; Podolsky & Schwed 1948; Born & Peng 1944) have proposed modifications of the framework of current quantum theory, some of which are successful in removing infinities from terms in the perturbation series which diverge in the orthodox development. The author has experimented with such modifications, and convinced himself that they are all expressible by a change of the Lagrangian function which generates the interacting fields, or alternatively by a change in the definition of the momentum and energy operators of the interacting particles, of such a kind that the orthodox theory is valid for low energies, but only approximately so as the energy is increased. Any number of non-divergent theories can be formulated in this way, and what is needed is some new physical principle to decide which of these is correct. As no principle of this kind has yet won universal acceptance, the situation which has evolved along these lines cannot be regarded as very satisfactory.

Before proceeding further in this direction, it would seem eminently worth while to re-examine the other alternative mentioned near the beginning, that the technique of the current perturbation theory is inadequate to deal with the problem. Peng (1946) has shown that some improvement is noticeable if one takes account of the degeneracy which is an essential feature of most interaction problems; this improvement does not, however, extend to the crucial example of the interaction between an electron and the electromagnetic field. Dirac (1947*a*) and Pauli (1947) seem to recognize the possibility that the perturbation expansion may be at least partly at fault, though they have no great hope of the required correction being made. In a recent paper (Born & Green 1948, appendix I), the author had occasion to develop an exact solution of the general perturbation problem, and the question at once arises whether use cannot be made of this solution to determine finally whether the difficulty is a purely mathematical one. The principal object of this paper is to answer this question as far as possible, by proceeding from much more exact formulae for the cross-sections and self-energies than those which are usually accepted. In this formalism, the cross-sections are *necessarily* convergent; but, although the self-energies may frequently converge where the usual perturbation series diverges, they are not finite of necessity, but each example has to be considered separately. The example which is crucial and at the same time of greatest physical interest concerns the interaction of the electron and photon fields. There is some hope that a determination of the self-energy of the electron, regarded as the source of its rest-mass, may afford the first acceptable theoretical value of the fine-structure constant; also it would greatly strengthen the present rather insecure theory of the Lamb-Rutherford shift (Bethe 1947). One may omit from consideration the electrostatic self-energy; the elaborate λ-limiting process reproduces, as it should, merely the Coulomb energy $\sum_{k<l} e_k e_l / r_{kl}$ which was originally inserted into the Hamiltonian; so that it is obviously better regarded as part of the energy of the unperturbed system. The transverse self-energy turns out to be rigorously divergent as long as the orthodox theory is employed. However, it is reduced to zero through the introduction of Dirac's negative-energy quanta (Dirac 1942; Pauli 1943) without the λ-limiting process, which is therefore completely eliminated.

Whether the introduction of negative-energy photons can be regarded as a satisfactory way out of the difficulties of quantum electrodynamics is still debatable; but there is included in the present paper a method of second quantization which makes it a very much more natural procedure than was originally apparent. A feature of the method employed throughout is the consistent use of the density matrix which the author has already found convenient in other connexions (Born & Green 1947; Green 1948). This requires a complete reformulation of the current field theory and method of second quantization, which are applicable only to wave vectors, thus leading in a very natural way to the appearance of the negative-energy states, the theory of which is much simplified.

2. THE GENERAL PERTURBATION THEORY

Before proceeding to the derivation of exact formulae for scattering cross-sections and self-energies, a method of obtaining an exact solution of the general perturbation problem must be given, substantially equivalent though superficially dissimilar to that developed by the author in another place (Born & Green 1948; appendix I). The work is much facilitated by the introduction of a new notation, in which operators of the form $A^m B A^n$ are denoted by $A^m A'^n B$, the prime indicating that the factor A^n should be written after the operand B when the term is rearranged in correct order. The notation is suggestive of the form $A^m A'^n B(A, A')$ of the corresponding representative when A is diagonal and $B(A, A')$ is the matrix element of B, but is used irrespective of representation. Thus an expression of the form $f(A, A')\,B$, where $f(A, A')$ is any function of A and A', means an operator whose matrix elements are $f(A, A')\,B(A, A')$ in a representation where A is diagonal; the matrix elements in any other representation can be obtained by making the appropriate unitary transformation. In this way the operator $f(A, A')\,B$ can be defined even when $f(A, A')$ is not an integral function. If $f(A, A')$ is a function with a singularity at $A = A'$, like $1/(A'-A)$, it is, however, necessary to avoid ambiguity that the diagonal elements $B(A, A)$ of B should vanish identically in the A-representation.

The representation actually adopted is one in which a set of variables, denoted collectively by the symbol λ, is diagonal, comprising the numbers of particles of the different types considered, together with their momenta and certain auxiliary variables, such as the sign of the energy state of an electron and the direction of polarization of a photon.

If H is the total Hamiltonian energy of the system, including interactions, the density matrix ρ representing the entire system will satisfy

$$i\hbar\frac{\partial\rho}{\partial t} = H\rho - \rho H = (H - H')\rho. \tag{2·1}$$

One can split H into two parts $\quad H = E_1 + V, \qquad (2·2)$

where E_1 is diagonal, and therefore expressible in terms of the representation variables λ, and V has vanishing diagonal elements. E_1 will generally represent the total energy of the system without interaction, and V the interaction energy. After the substitutions

$$\rho = \mathrm{e}^{E_1 t/i\hbar}\,\tilde{\rho}_1\,\mathrm{e}^{-E_1 t/i\hbar} = \mathrm{e}^{(E_1 - E_1')t/i\hbar}\,\tilde{\rho}_1, \quad V = \mathrm{e}^{(E_1 - E_1')t/i\hbar}\,\tilde{v}_1, \tag{2·3}$$

the equation (2·1) reduces to

$$i\hbar\frac{\partial\tilde{\rho}_1}{\partial t} = \tilde{v}_1\tilde{\rho}_1 - \tilde{\rho}_1\tilde{v}_1 = (\tilde{v}_1 - \tilde{v}_1')\,\tilde{\rho}_1. \tag{2·4}$$

This suggests the substitution

$$\tilde{\rho}_1 = \mathrm{e}^{(\tilde{u}_1 - \tilde{u}_1')/i\hbar}\,\rho_2, \quad \tilde{u}_1 = \int_0^t \tilde{v}_1\,dt. \tag{2·5}$$

Now, since

$$\left.\begin{aligned} \frac{\partial}{\partial t}\tilde{u}_1^n &= (\tilde{u}_1^{n-1} + \tilde{u}_1^{n-2}\tilde{u}_1' + \ldots + \tilde{u}_1'^{n-1})\,\tilde{v}_1 = \left(\frac{\tilde{u}_1^n - \tilde{u}_1'^n}{\tilde{u}_1 - \tilde{u}_1'}\right)\tilde{v}_1, \\ \frac{\partial}{\partial t}\mathrm{e}^{\tilde{u}_1/i\hbar} &= \mathrm{e}^{\tilde{u}_1/i\hbar}\left(\frac{\mathrm{e}^{(\tilde{u}_1' - \tilde{u}_1)/i\hbar} - 1}{\tilde{u}_1' - \tilde{u}_1}\right)\tilde{v}_1, \end{aligned}\right\} \tag{2·6}$$

one obtains

$$i\hbar\frac{\partial\rho_2}{\partial t} = H_2\rho_2 - \rho_2 H_2 = (H_2 - H_2')\,\rho_2, \tag{2·7}$$

where
$$H_2 = \left\{\left(1-\frac{i\hbar}{\tilde{u}_1'-\tilde{u}_1}\right)\mathrm{e}^{(\tilde{u}_1'-\tilde{u}_1)/i\hbar}+\frac{i\hbar}{\tilde{u}_1'-\tilde{u}_1}\right\}\tilde{v}_1$$
$$= \left\{\frac{1}{2!}\left(\frac{\tilde{u}_1'-\tilde{u}_1}{i\hbar}\right)+\frac{2}{3!}\left(\frac{\tilde{u}_1'-\tilde{u}_1}{i\hbar}\right)^2+\frac{3}{4!}\left(\frac{\tilde{u}_1'-\tilde{u}_1}{i\hbar}\right)^3+\ldots\right\}\tilde{v}_1. \qquad (2{\cdot}8)$$

In the same way, H_2 can be split into a diagonal part E_2 and a part v_2 with vanishing diagonal elements
$$H_2 = E_2+v_2, \qquad (2{\cdot}9)$$
and by making the substitutions
$$\rho_2 = \mathrm{e}^{(E_2-E_2')t/i\hbar}\,\tilde{\rho}_2, \quad v_2 = \mathrm{e}^{(E_2-E_2')t/i\hbar}\,\tilde{v}_2, \qquad (2{\cdot}10)$$
and
$$\tilde{\rho}_2 = \mathrm{e}^{(\tilde{u}_2-\tilde{u}_2')/i\hbar}\,\rho_3, \quad \tilde{u}_2 = \int_0^t \tilde{v}_2\,dt, \qquad (2{\cdot}11)$$
the equation (2·7) is reduced to
$$\left.\begin{aligned} i\hbar\frac{\partial\rho_3}{\partial t} &= (H_3-H_3')\,\rho_3,\\ H_3 &= \left\{\left(1-\frac{i\hbar}{\tilde{u}_2'-\tilde{u}_2}\right)\mathrm{e}^{(\tilde{u}_2'-\tilde{u}_2)/i\hbar}+\frac{i\hbar}{\tilde{u}_2'-\tilde{u}_2}\right\}\tilde{v}_2.\end{aligned}\right\} \qquad (2{\cdot}12)$$

Since H_2 is quadratic in v_1 and H_3 is quartic, etc., this process is rapidly convergent if v_1 is small in relation to E_1; and the exact value of ρ at time t is
$$\rho = \mathrm{e}^{E_1t/i\hbar}\,\mathrm{e}^{\tilde{u}_1/i\hbar}\,\mathrm{e}^{E_2t/i\hbar}\,\mathrm{e}^{\tilde{u}_2/i\hbar}\ldots\tilde{\rho}_0\ldots\mathrm{e}^{-\tilde{u}_2/i\hbar}\,\mathrm{e}^{-E_2t/i\hbar}\,\mathrm{e}^{-\tilde{u}_1/i\hbar}\,\mathrm{e}^{-E_1t/i\hbar}, \qquad (2{\cdot}13)$$
where $\tilde{\rho}_0$ is the (arbitrary) value at time $t = 0$. This expression reveals an important advantage of the present method, that ρ is related to $\tilde{\rho}_0$ by a unitary transformation, and the trace of these two operators is therefore the same; hence, if (2·13) is correctly interpreted, cross-sections for scattering derived from it cannot diverge, and the same is true also of the intermediate approximations ρ_2, ρ_3, etc.

The equation (2·13) may be rewritten
$$\rho = \mathrm{e}^{u_1/i\hbar}\,\mathrm{e}^{u_2/i\hbar}\ldots\rho_0\ldots\mathrm{e}^{-u_2/i\hbar}\,\mathrm{e}^{-u_1/i\hbar}, \qquad (2{\cdot}14)$$
where
$$u_1 = \mathrm{e}^{E_1t/i\hbar}\,\tilde{u}_1\,\mathrm{e}^{-E_1t/i\hbar}, \quad u_2 = \mathrm{e}^{(E_1+E)_2t/i\hbar}\,\tilde{u}_2\,\mathrm{e}^{-(E_1+E_2)t/i\hbar}, \quad \text{etc.}, \qquad (2{\cdot}15)$$
and
$$\rho_0 = \mathrm{e}^{Et/i\hbar}\,\tilde{\rho}_0\,\mathrm{e}^{-Et/i\hbar}, \quad E = E_1+E_2+\ldots. \qquad (2{\cdot}16)$$

It may be inferred from (2·16) that E is the total energy of the system in diagonal form, including the interaction energies and self-energies of the system. To obtain cross-sections, the usual procedure is to substitute a diagonal value for $\tilde{\rho}_0$, representing a pure state, and to seek the asymptotic value of ρ for large values of t. A more direct method is to find the asymptotic value of $\dot{\rho} = (H\rho-\rho H)/i\hbar$, since it is the rate of change of the diagonal elements of ρ which is actually required. In the calculation of these asymptotic values, one is effectively making the transition from the above time-dependent theory to the corresponding theory of stationary states.

On the assumption that V is independent of time, one obtains from (2·5) and (2·3)
$$\tilde{u}_1 = \int_0^t \mathrm{e}^{(E_1'-E_1)t/i\hbar}\,V\,dt = \left(\frac{\mathrm{e}^{(E_1'-E_1)t/i\hbar}-1}{E_1'-E_1}\right)i\hbar V,$$

so, according to (2·15), $$u_1 = \left(\frac{1-\mathrm{e}^{(E_1-E_1')t/i\hbar}}{E_1'-E_1}\right) i\hbar V. \tag{2·17}$$

Now, if ϵ is chosen sufficiently small, but ϵt nevertheless sufficiently large, the expression $\frac{1-\mathrm{e}^{(E_1-E_1')t/i\hbar}}{E_1'-E_1}$ may, in integrals with respect to E_1', be replaced with negligible error by $1/(E_1'-E_1)$ for $|E_1'-E_1|>\epsilon$ and by $-i\pi\delta(E_1'-E_1)$ for $|E_1'-E_1|<\epsilon$. Hence, one has

$$U_1 = \lim_{t\to\infty} u_1 = \left\{\frac{i\hbar}{E_1'-E_1}+\pi\hbar\delta(E_1'-E_1)\right\} V, \tag{2·18}$$

which satisfies $$(U_1-U_1')\,E_1 = (E_1'-E_1)\,U_1 = i\hbar V. \tag{2·19}$$

The matrix representative of U_1 will be recognized as the transformation matrix which appears in all theories of collision (cf. Heisenberg 1943). Let

$$\mathrm{e}^{-U_1/i\hbar} H\, \mathrm{e}^{U_1/i\hbar} = E_1+E_2+V_2, \tag{2·20}$$

where E_2 is diagonal and V_2 has vanishing diagonal elements. Writing $H = E_1+V$ in (2·20) and using (2·19) to eliminate E_1, one obtains

$$E_2+V_2 = \left\{\left(1-\frac{i\hbar}{U_1'-U_1}\right)\mathrm{e}^{(U_1'-U_1)/i\hbar}+\frac{i\hbar}{U_1'-U_1}\right\} V, \tag{2·21}$$

so V_2 is the asymptotic value of v_2, and

$$U_2 = \left[\frac{i\hbar}{(E_1'+E_2')-(E_1+E_2)}+\pi\hbar\delta\{(E_1'+E_2')-(E_1+E_2)\}\right] V_2 \tag{2·22}$$

of u_2. The development may be continued by writing

$$\mathrm{e}^{-U_2/i\hbar}\,\mathrm{e}^{-U_1/i\hbar} H\, \mathrm{e}^{U_1/i\hbar}\,\mathrm{e}^{U_2/i\hbar} = E_1+E_2+E_3+V_3, \tag{2·23}$$

etc. The asymptotic value of ρ is

$$P = \mathrm{e}^{U_1/i\hbar}\,\mathrm{e}^{U_2/i\hbar}\ldots\rho_0\ldots\mathrm{e}^{-U_2/i\hbar}\,\mathrm{e}^{-U_1/i\hbar}, \tag{2·24}$$

and since ρ_0, defined by (2·16), satisfies

$$i\hbar\dot{\rho}_0 = (E-E')\,\rho_0, \tag{2·25}$$

and further $$H\,\mathrm{e}^{U_1/i\hbar}\,\mathrm{e}^{U_2/i\hbar}\ldots = \mathrm{e}^{U_1/i\hbar}\,\mathrm{e}^{U_2/i\hbar}\ldots E, \tag{2·26}$$

this value P actually satisfies the equation (2·1).

3. THE FORMULAE FOR CROSS-SECTIONS AND SELF-ENERGIES

In specific applications, the initial density matrix $\tilde{\rho}_0$ is taken diagonal (usually with a single non-vanishing diagonal element), and it might then appear that ρ_0 and P, defined in (2·24), ought to be independent of time. This is not so, however, as may be seen from the following consideration. The diagonal elements of P contain terms with a factor of the form $\left|\frac{i\hbar}{E'-E}+\pi\hbar\delta(E'-E)\right|^2$, and to interpret this singularity correctly it is necessary to return to the expression

$$\left|\frac{i\hbar(1-\mathrm{e}^{(E-E')t/i\hbar})}{E'-E}\right|^2 \quad (t \text{ large}),$$

from which it was derived. One obtains

$$\begin{aligned}\frac{d}{dt}\left|\frac{i\hbar(1-\mathrm{e}^{(E-E')t/i\hbar})}{E'-E}\right|^2 &= \frac{2\hbar}{E-E'}\sin(E-E'')\,t/\hbar \quad (t \text{ large})\\ &= 2\pi\hbar\delta(E-E'),\end{aligned}$$

so that one has

$$\left|\frac{i\hbar}{E-E'}+\pi\hbar\delta(E-E')\right|^2 = 2\pi\hbar t\delta(E'-E). \tag{3·1}$$

Cross-sections for scattering are determined by the asymptotic rate of change of the diagonal elements of the density matrix, and require the calculation of the limiting form of $\dot{\rho} = H\rho-\rho H$, or, since the diagonal elements of $E_1\rho-\rho E_1$ vanish, of $V\rho-\rho V$. They are therefore given accurately by the diagonal elements of

$$\dot{P} = \frac{VP-PV}{i\hbar}, \tag{3·2}$$

which is time independent. In practice one may substitute to a sufficient degree of accuracy

$$\dot{P}_1 = (i\hbar)^{-1}(V\,\mathrm{e}^{U_1/i\hbar}\rho_0\,\mathrm{e}^{-U_1/i\hbar}-\mathrm{e}^{U_1/i\hbar}\rho_0\,\mathrm{e}^{-U_1/i\hbar}\,V). \tag{3·3}$$

In the calculation of self-energies one naturally limits consideration to one type *of particle in interaction with its associated field. The self-energy is then contained* in $E-E_1$, and is given sufficiently nearly by the diagonal elements of E_2, defined by (2·20). Discarding V_2, which has no diagonal elements, on the right-hand side of this equation, and replacing H by E_1 to a similar degree of approximation on the left, one obtains

$$E_2 \sim \mathrm{e}^{-U_1/i\hbar}\,E_1\,\mathrm{e}^{U_1/i\hbar}-E_1. \tag{3·4}$$

For the explicit evaluation of (3·3) and (3·4), it is necessary to know the matrix elements of $\mathrm{e}^{U_1/i\hbar}$. Let U be an eigenvalue, and $\psi(\lambda, U)$ the corresponding normalized eigenvector, of the operator U_1, so that

$$\left.\begin{aligned}\sum_{\lambda'} U_1(\lambda,\lambda')\,\psi(\lambda',U) &= U\psi(\lambda,U),\\ \sum_{\lambda}\psi^*(\lambda,U)\,\psi(\lambda,U') &= \delta(U-U').\end{aligned}\right\} \tag{3·5}$$

Then

$$\mathrm{e}^{U_1/i\hbar}(\lambda,\lambda') = \int\psi(\lambda,U)\,\mathrm{e}^{U/i\hbar}\,\psi^*(\lambda',U)\,dU, \tag{3·6}$$

and one obtains

$$\left.\begin{aligned}\hbar\dot{P}_1(\lambda',\lambda') &= 2\mathscr{I}[\sum_{\lambda''}\{V(\lambda',\lambda'')\,\mathrm{e}^{U_1/i\hbar}(\lambda'',\lambda)\}\,\mathrm{e}^{-U_1/i\hbar}(\lambda,\lambda')],\\ E_2(\lambda) &= -E_1+\sum_{\lambda}|\,\mathrm{e}^{U_1/i\hbar}(\lambda,\lambda')\,|^2\,E_1',\end{aligned}\right\} \tag{3·7}$$

where λ represents the initial state. It is obvious that these expressions will converge in a variety of circumstances where the ordinary perturbation series, obtained by expanding the exponential $e^{U/i\hbar} = 1 + U/i\hbar - U^2/\hbar^2 + \dots$ in (3·6), diverges. Indeed, only in the most favourable conditions, when $\psi(\lambda, U)$ tends to zero as $U \to \infty$ more quickly than any power of U^{-1}, can the ordinary perturbation series be expected to converge. The cross-sections obtained from (3·7), on the other hand, are necessarily convergent on account of the unitary character of $e^{U_1/i\hbar}$, whilst the self-energy can diverge only on account of the summation over λ'.

4. THE FIELD THEORY OF THE DENSITY MATRIX

Before the second quantization of the density matrix ρ can be considered, it is necessary to devote some attention to the logically prior problem presented by a relativistic Lagrangian formulation of field theory in which the density matrix replaces the wave vector with which the customary theory operates. The notation of general relativity theory will be employed, but the metric tensor $g_{\alpha\beta}$ will be supposed to have the Galilean components $g_{11} = g_{22} = g_{33} = -1$, $g_{44} = 1$, $g_{\alpha\beta} = 0$, $\alpha \neq \beta$. One may further conveniently make use of the relativistic form of the density matrix $\rho(x^\alpha, x^{\alpha'})$, $x^\alpha \equiv (\mathrm{x}, ct)$ in the co-ordinate representation, discussed by the author (Green 1949) elsewhere; the ordinary density matrix is obtained from it by writing $x^{4'} = x^4$. It is sufficient for the present purpose to consider only Lagrangians linear in the density matrix, though any number of space and time derivatives will be allowed; some of the corresponding theory in terms of wave vectors is well known (cf. Chang 1946; de Wet 1948), but is more complicated than that given here (vide appendix). For the purpose of the present theory, which will be developed quite independently, a Lagrangian operator $\mathscr{L}$ is formed, expressed as the sum of terms of the form $c_{mn} p_1^{m_1} p_2^{m_2} p_3^{m_3} p_4^{m_4} p_1'^{n_1} p_2'^{n_2} p_3'^{n_3} p_4'^{n_4} \rho$, where p_α is the momentum operator with the space representative $-i\hbar \dfrac{\partial}{\partial x^\alpha} \prod\limits_\beta \delta(x^\beta - x^{\beta'})$, and the prime attached to p'_α signifies, in accordance with the notation introduced in § 2, that this factor is normally written after the operand ρ. Then one may write

$$\mathscr{L} = F(p_\alpha, p'_\alpha)\rho, \tag{4·1}$$

where F is a given function of the four-vectors p_α and p'_α, which may clearly be replaced by $-i\hbar(\partial/\partial x^\alpha)$ and $i\hbar(\partial/\partial x^{\alpha'})$ respectively in the co-ordinate representation. The Lagrangian density L is obtained from $\mathscr{L}$ by writing $x^{\alpha'} = x^\alpha$, and the action Λ by taking the trace $\mathrm{X}(\mathscr{L})$ of that operator.

It is convenient to define the following:

$$\pi_\alpha = p'_\alpha - p_\alpha, \quad \pi'_\alpha = p_\alpha - p'_\alpha, \quad \pi_\alpha + \pi'_\alpha = 0; \tag{4·2}$$

$$\left.\begin{aligned} G^{\alpha\beta} &= \frac{\partial^2}{\partial p_\alpha \partial p_\beta}\left(\frac{1}{2!}F + \frac{\pi_\gamma}{3!}\frac{\partial F}{\partial p_\gamma} + \frac{\pi_\gamma \pi_\delta}{4!}\frac{\partial^2 F}{\partial p_\gamma \partial p_\delta} + \dots\right), \\ G^{\alpha\beta'} &= \frac{\partial^2}{\partial p'_\alpha \partial p'_\beta}\left(\frac{1}{2!}F + \frac{\pi'_\gamma}{3!}\frac{\partial F}{\partial p'_\gamma} + \frac{\pi'_\gamma \pi'_\delta}{4!}\frac{\partial^2 F}{\partial p'_\gamma \partial p'_\delta} + \dots\right); \end{aligned}\right\} \tag{4·3}$$

$$G^\alpha = \frac{\partial F}{\partial p_\alpha} + \pi_\beta G^{\alpha\beta}, \quad G^{\alpha'} = \frac{\partial F}{\partial p'_\alpha} + \pi'_\beta G^{\alpha\beta'}; \tag{4·4}$$

so that, by Taylor's theorem,

$$F(p_\alpha, p_\alpha) = F(p_\alpha, p'_\alpha) + \pi'_\alpha G^{\alpha\prime}, \quad F(p'_\alpha, p'_\alpha) = F(p_\alpha, p'_\alpha) + \pi_\alpha G^\alpha. \tag{4·5}$$

Euler's field equations can be obtained from the variational principle of 'least action' in the following way:

$$\delta\Lambda = \mathrm{X}\{F(p_\alpha, p'_\alpha)\,\delta\rho\} = \mathrm{X}\{F(p_\alpha, p_\alpha)\,\delta\rho\} - \mathrm{X}(\pi'_\alpha G^{\alpha\prime}\delta\rho),$$

according to (4·5). The second term on the right-hand side is

$$i\hbar\int \frac{\partial}{\partial x^\beta}\{(G^{\beta\prime}\delta\rho)\,(x^\alpha, x^\alpha)\}\,dx$$

and reduces to a three-dimensional surface integral which vanishes if $\delta\rho$ is supposed to vanish at the boundary of the four-dimensional region of integration. Hence one obtains the field equations

$$F(p_\alpha, p_\alpha)\,\rho \equiv \mathscr{L} + \pi'_\alpha G^{\alpha\prime}\rho = 0, \quad F(p'_\alpha, p'_\alpha)\,\rho \equiv \mathscr{L} + \pi_\alpha G^\alpha\rho = 0, \tag{4·6}$$

the second being the conjugate of the first.

If now one defines the tensor operator

$$\mathscr{L}^{\alpha\beta} = -\mathscr{L}g^{\alpha\beta} + (p^\beta G^\alpha + p^{\beta\prime}G^{\alpha\prime})\,\rho, \tag{4·7}$$

the trace of which is the energy-momentum tensor, it follows from (4·6) that

$$\pi_\alpha \mathscr{L}^{\alpha\beta} = 0, \tag{4·8}$$

which is the operational form of the equation of conservation of energy and momentum.

For the symmetrization of $\mathscr{L}^{\alpha\beta}$ one needs to make use of the Lorentz invariance of $\mathscr{L}$; the corresponding theory in terms of wave vectors is given for comparison in the appendix. Making the infinitesimal Lorentz transformation

$$x^\alpha \to x^\alpha + \omega^\alpha_{\cdot\beta}x^\beta \quad (\omega_{\alpha\beta} = g_{\alpha\gamma}\omega^\gamma_{\cdot\beta} = -\omega_{\beta\alpha}),$$

$$p_\alpha \to p_\alpha - \omega^\beta_{\cdot\alpha}p_\beta,$$

let the (anti-Hermitian) operator $S^\alpha_{\cdot\beta}$ $(S_{\alpha\beta} = g_{\alpha\gamma}S^\gamma_{\cdot\beta} = -S_{\beta\alpha})$ be defined by

$$\rho \to \rho + \omega^\beta_{\cdot\alpha}(S^\alpha_{\cdot\beta}\rho - \rho S^\alpha_{\cdot\beta}) = \exp(\omega^\beta_{\cdot\alpha}S^\alpha_{\cdot\beta})\,\rho\,\exp(-\omega^\beta_{\cdot\alpha}S^\alpha_{\cdot\beta}).$$

Then

$$\mathscr{L} \to \mathscr{L} + \omega^\beta_{\cdot\alpha}\left\{(S^\alpha_{\cdot\beta} - S^{\alpha\prime}_{\cdot\beta})\,\mathscr{L} - \left(p_\beta\frac{\partial F}{\partial p_\alpha} + p'_\beta\frac{\partial F}{\partial p'_\alpha}\right)\rho\right\}, \tag{4·9}$$

so that, if $\mathscr{L}$ is Lorentz-invariant,

$$(S^{\alpha\beta} - S^{\alpha\beta\prime})\,\mathscr{L} = \frac{1}{2}\left(p^\beta\frac{\partial F}{\partial p_\alpha} + p^{\beta\prime}\frac{\partial F}{\partial p'_\alpha} - p^\alpha\frac{\partial F}{\partial p_\beta} - p^{\alpha\prime}\frac{\partial F}{\partial p'_\beta}\right)\rho. \tag{4·10}$$

Now, let $\mathscr{L}^{\alpha\beta,\gamma} = -\mathscr{L}^{\beta\alpha,\gamma}$ be defined by

$$\mathscr{L}^{\alpha\beta,\gamma} = (S^{\alpha\beta}G^{\gamma\prime} + S^{\alpha\beta\prime}G^\gamma)\,\rho + \tfrac{1}{2}(p^\beta G^{\alpha\gamma} - p^{\beta\prime}G^{\alpha\gamma\prime} - p^\alpha G^{\beta\gamma} + p^{\alpha\prime}G^{\beta\gamma\prime})\,\rho, \tag{4·11}$$

and form the product of this expression with π_γ; the first term reduces to $(S^{\alpha\beta} - S^{\alpha\beta\prime})\,\mathscr{L}$ by virtue of (4·6), and the second to

$$\tfrac{1}{2}(p^\beta G^\alpha + p^{\beta\prime}G^{\alpha\prime} - p^\alpha G^\beta - p^{\alpha\prime}G^{\beta\prime})\,\rho - \frac{1}{2}\left(p^\beta\frac{\partial F}{\partial p_\alpha} + p^{\beta\prime}\frac{\partial F}{\partial p'_\alpha} - p^\alpha\frac{\partial F}{\partial p_\beta} - p^{\alpha\prime}\frac{\partial F}{\partial p'_\beta}\right)\rho$$

by virtue of (4·4); hence, making use of (4·10), one obtains

$$\pi_\gamma \mathcal{L}^{\alpha\beta,\gamma} = \tfrac{1}{2}(\mathcal{L}^{\alpha\beta} - \mathcal{L}^{\beta\alpha}). \tag{4·12}$$

It follows from this that the angular momentum-tensor operator,

$$\mathcal{M}^{\alpha\beta,\gamma} = \mathcal{L}^{\alpha\beta,\gamma} + (2i\hbar)^{-1}(x^\alpha \mathcal{L}^{\gamma\beta} - x^\beta \mathcal{L}^{\gamma\alpha}) \tag{4·13}$$

satisfies the conservation law $\pi_\gamma \mathcal{M}^{\alpha\beta,\gamma} = 0$, and that the symmetrical energy momentum tensor operator, defined by

$$\mathcal{T}^{\alpha\beta} = \tfrac{1}{2}(\mathcal{L}^{\alpha\beta} + \mathcal{L}^{\beta\alpha}) + \pi_\gamma(\mathcal{L}^{\gamma\alpha,\beta} + \mathcal{L}^{\gamma\beta,\alpha}) \tag{4·14}$$

also satisfies the conservation law.

The Lagrangian expressed in terms of the density matrix is automatically gauge-invariant; it might be anticipated, therefore, that no new assumption would be required to demonstrate the conservation of charge, and from the expression

$$\mathcal{S}^\alpha = (G^\alpha + G^{\alpha'})\rho \tag{4·15}$$

for the charge-current vector, it is indeed obvious that, by virtue of the field equations (4·6), $\pi_\alpha \mathcal{S}^\alpha = 0$.

5. THE SECOND QUANTIZATION OF THE DENSITY MATRIX

The usual method of second quantization proceeds from the commutation rules satisfied by the wave vector, which are determined fundamentally by the requirement that they should lead to the particular aspect of matter. This method, besides being out of keeping with the fundamental role of the density matrix, has an important defect in that it excludes negative expectation values of the density matrix. While the precise meaning of such negative expectation values requires additional clarification, the necessity for their introduction first became apparent in positron theory (cf. Heisenberg 1934), where it found expression in a correspondence principle of a pragmatic nature, with the sole function of making the energy positive definite. Dirac (1942) has devised an elaborate formalism designed to extend the concept of negative expectation values to particles satisfying Bose statistics, and made use of it to alleviate, without actually in itself removing, the divergences of orthodox theory. Though the principal object here is simply the improvement of the rigour of the orthodox theory, it will be shown that the necessity for such complications disappears when second quantization is applied to the density matrix instead of to the wave vector. This theory is now developed in an heuristic way.

Introducing as a temporary measure the wave function $\psi(n)$ representing a state in which n particles, of a given kind and with specified momentum and other auxiliary variables, are present, let ζ be the creation operator for this kind of particle, satisfying

$$\zeta\psi(n) = \psi(n+1). \tag{5·1}$$

The operator ζ is, of course, diagonal in the momentum, and, as both $\psi(n)$ and $\psi(n+1)$ are normalized, must satisfy $\zeta^*\zeta = 1$; it is a unitary operator, though it would be incorrect to identify it with the quasi-unitary operator occurring in the theory of the harmonic oscillator (see Dirac 1947, § 34), where n is limited to positive values. The density matrix ρ has the property

$$\rho\psi(n) = n\psi(n). \tag{5·2}$$

It is now proposed to formulate the commutation rule for ρ and ζ; this will depend on the range of values which may be assumed by n, and one has therefore to consider several different alternatives.

First, suppose n may be taken to have any integral value, positive or negative; then, for all values of n,

$$\begin{aligned} \rho\zeta\psi(n) &= \rho\psi(n+1) = (n+1)\,\psi(n+1) \\ &= \zeta(n+1)\,\psi(n) = \zeta(\rho+1)\,\psi(n); \end{aligned}$$

hence

$$\rho\zeta - \zeta\rho = \zeta \quad (\zeta^*\zeta = 1) \tag{5·3}$$

is the required commutation rule. The usual methods, applied to (5·3), lead to the following matrix representatives:

$$\rho_{nn'} = n\delta_{nn'}, \quad \zeta_{nn'} = \delta_{n,n'+1}. \tag{5·4}$$

Introducing a new operator σ by means of the relation

$$\rho = (\sigma\zeta^*)^2 = \sigma\sigma^\dagger, \quad \sigma^\dagger = \zeta^*\sigma\zeta^*, \tag{5·5}$$

the commutation rule (5·3) reduces to

$$\sigma^\dagger\sigma - \sigma\sigma^\dagger = 1, \tag{5·6}$$

since $\sigma^\dagger\sigma = \zeta^*\sigma\zeta^*\zeta\sigma^\dagger\zeta = \zeta^*\rho\zeta$. As, in (5·2) and (5·4), n may take negative values, this does not yet correspond to Bose statistics, but to what one might call Dirac statistics, in which negative expectation values of the density matrix are allowed. To obtain Bose statistics proper, one has to add the condition that ρ is positive definite; this is achieved by identifying $\sigma^\dagger$ with the Hermitian conjugate of σ, when (3·6) reduces to the usual commutation rule, and one obtains the representative $\sigma_{nn'} = n^{\frac{1}{2}}\,\delta_{nn'+1}$; further, if $i\hbar\dot{\sigma} = H\sigma$, (2·1) is satisfied on account of (5·5). One may suppress the positive expectation values in the same way by imposing the condition that $\sigma^\dagger$ should be the *negative* of the Hermitian conjugate of σ; then the representative is $\sigma_{nn'} = i\sqrt{(-n)}\,\delta_{n,n'-1}$ with n and n' taking only values which are not positive. In general, however, there is no need to separate the positive and negative expectation values in this artificial way, and one may write $\sigma_{nn'} = n^{\frac{1}{2}}\,\delta_{|n-n'|,1}\,\delta_{|n|-|n'|,1}$ for both positive and negative values of n and n'.

Next, suppose n may take only the values 0 and 1; then

$$\zeta\psi(0) = \psi(1), \quad \zeta^*\psi(1) = \psi(0).$$

If $\zeta\psi(1) = a\psi(0) + b\psi(1) = (a\zeta^* + b)\,\psi(1)$, then $a\zeta^* + b$ must be unitary, and it follows that either $a = 0$ and $|\,b\,| = 1$ or $b = 0$ and $|\,a\,| = 1$. Adopting the more symmetrical second alternative, which implies that ζ and ζ^* differ only by a numerical factor of modulus unity, one has instead of (5·1)

$$\zeta\psi(n) = \psi(1-n) \tag{5·7}$$

if one takes $\zeta = \zeta^*$, and

$$\rho\zeta\psi(n) = (1-n)\,\psi(1-n) = \zeta(1-\rho)\,\psi(n),$$

so that

$$\rho\zeta + \zeta\rho = \zeta \quad (\zeta^2 = 1). \tag{5·8}$$

The matrix representatives of ρ and ζ are

$$\rho = n\delta_{nn'}, \quad \zeta = \delta_{n,1-n'} \quad (n, n' = 0, 1). \tag{5·9}$$

This corresponds to Fermi statistics, but does not lead to a correct positron theory with positive definite energy without the addition of an awkward correspondence postulate. To avoid this, one may make use of the complementary theory in which n takes the values 0 and -1; one then obtains instead of (5·8),

$$\rho\zeta + \zeta\rho = -\zeta \tag{5·10}$$

for the negative energy states. The anti-commutation rules (5·8) and (5·10) may then be combined in the single formula

$$\rho\zeta + \zeta\rho = \zeta\tau, \tag{5·11}$$

where τ is a Pauli matrix with proper values ± 1; (5·11) is valid for electrons *and* positrons. τ can be related to the anti-commuting Dirac matrices $\boldsymbol{\alpha}$, β by means of $(p^2 + m^2c^2)^{\frac{1}{2}}\tau = \mathbf{p}\,.\,\boldsymbol{\alpha} + mc\beta$. One may interpret Dirac's theory of holes in the sense that the negative expectation value -1 signifies the *absence* of a particle which is present, though unobservable, *in vacuo*. This being so, one would expect the same, to apply to the negative expectation values which arise in the generalization of Bose statistics; it is true that there is then the further conceptual difficulty that the number of Bose particles which can occupy a negative energy state is unlimited, and the state of the vacuum is therefore left quite arbitrary; however, as there seems to be no way of distinguishing between a photon in a positive energy state and a 'gap' in a negative energy state, this does not lead to insuperable difficulties in practice.

6. THE ELECTRON AND PHOTON FIELDS IN INTERACTION

The general theory of the preceding sections can now be applied to the special example of the electron and electromagnetic fields in interaction, where the divergence difficulties have hitherto been most troublesome. No deviation from strictly orthodox theory need yet be considered; thus, although positron theory is admitted, the photon field may be supposed quantized strictly in accordance with Bose statistics, and the changes occasioned by the introduction of negative energy photons will be left for notice at the end. The vector potential of the electromagnetic field is represented in Fourier series by

$$\mathbf{A}(\mathbf{x}) = \left(\frac{2\pi c}{\Omega}\right)^{\frac{1}{2}} \hbar \sum_r k_r^{-\frac{1}{2}} \mathbf{a}_r (\sigma_r \, \mathrm{e}^{\iota \mathbf{k}_r \cdot \mathbf{x}/\hbar} + \sigma_r^\dagger \, \mathrm{e}^{-\iota \mathbf{k}_r \cdot \mathbf{x}/\hbar}), \tag{6·1}$$

where the suffix r refers to photons with momentum $\mathbf{k}_r$ and direction of polarization $\mathbf{a}_r$, σ_r is the corresponding operator defined by (5·5), and Ω is the volume of the real space considered. In the λ-representation, the interaction energy $V = e\boldsymbol{\alpha}.\mathbf{A}$ has therefore the representative

$$V(\lambda,\lambda') = \left(\frac{2\pi c}{\Omega}\right)^{\frac{1}{2}} e\hbar \sum_r k_r^{-\frac{1}{2}} \mathbf{a}_r.\boldsymbol{\alpha}(n_r^{\frac{1}{2}}\delta_{n,n'+1_r}\delta_{\mathbf{p},\mathbf{p}'-\mathbf{k}_r} + n_r'^{\frac{1}{2}}\delta_{n,n'-1,}\delta_{\mathbf{p},\mathbf{p}'+\mathbf{k}_r}), \tag{6·2}$$

where $\delta_{n,n'+1_r} = \delta_{n_r,n_r'+1}\prod_s \delta_{n_s n_s'}$. Also

$$E_1(\lambda,\lambda') = c(\sum_r n_r k_r + \mathbf{p}.\boldsymbol{\alpha} + mc\beta)\,\delta_{nn'}\delta_{\mathbf{pp}'}, \tag{6·3}$$

where $\delta_{nn'} = \prod_r \delta_{n_r n_r'}$. As in § 5, $\boldsymbol{\alpha}, \beta$ constitute a set of anti-commuting Dirac matrices. The non-vanishing elements of V are

$$\left.\begin{aligned} V(\mathbf{p},n;\mathbf{p}+\mathbf{k}_r,n-1_r) &= \left(\frac{2\pi c n_r}{\Omega k_r}\right)^{\frac{1}{2}} e\hbar\mathbf{a}_r.\boldsymbol{\alpha}, \\ V(\mathbf{p},n;\mathbf{p}-\mathbf{k}_r,n+1_r) &= \left(\frac{2\pi c \overline{n_r+1}}{\Omega k_r}\right)^{\frac{1}{2}} e\hbar\mathbf{a}_r.\boldsymbol{\alpha}, \end{aligned}\right\} \tag{6·4}$$

and as the difference

$$\frac{E_1 - E_1'}{c} = \pm k_r \pm (m^2c^2 + p^2)^{\frac{1}{2}} \pm \{m^2c^2 + (\mathbf{p} \pm \mathbf{k}_r)^2\}^{\frac{1}{2}}$$

can in no circumstances vanish, the term containing $\delta(E_1 - E_1')$ may be omitted from (2·18), and the non-vanishing elements of U_1 are given by

$$U_1(\mathbf{p},n;\mathbf{p}+\mathbf{k}_r,n-1_r) = \left(\frac{2\pi n_r}{c\Omega k_r}\right)^{\frac{1}{2}} e\hbar \frac{i\hbar}{E_1'/c - mc\beta - \mathbf{p}.\boldsymbol{\alpha} - \sum_s n_s k_s}.\mathbf{a}_r.\boldsymbol{\alpha}.$$

In accordance with the notation introduced in § 2, E_1' may be replaced by its value $mc\beta + (\mathbf{p}+\mathbf{k}_r).\boldsymbol{\alpha} + \sum_s (n_s - \delta_{rs})\,k_s$ only after the term in which it occurs has been removed to a position following the operand $\mathbf{a}_r.\boldsymbol{\alpha} = \dfrac{1}{\mathbf{a}_r.\boldsymbol{\alpha}}$. Use is made in effecting this removal of the fact that $\mathbf{a}_r.\boldsymbol{\alpha}$ and $(\mathbf{p} - \mathbf{p}.\mathbf{a}_r\mathbf{a}_r).\boldsymbol{\alpha}$ anti-commute, since $\mathbf{a}_r$ and $\mathbf{p} - \mathbf{p}.\mathbf{a}_r\mathbf{a}_r$ are perpendicular; also $\mathbf{a}_r.\boldsymbol{\alpha}$ and β anti-commute. Hence

$$\begin{aligned} U_1(\mathbf{p},n;\mathbf{p}+\mathbf{k}_r,n-1_r) &= \left(\frac{2\pi n_r}{c\Omega k_r}\right)^{\frac{1}{2}} \frac{e\hbar}{\mathbf{a}_r.\boldsymbol{\alpha}} \frac{i\hbar}{E_1'/c + mc\beta + (\mathbf{p} - 2\mathbf{p}.\mathbf{a}_r\mathbf{a}_r).\boldsymbol{\alpha} - \sum_s n_s k_s} \\ &= \left(\frac{2\pi n_r}{c\Omega k_r}\right)^{\frac{1}{2}} \frac{e\hbar}{\mathbf{a}_r.\boldsymbol{\alpha}} \frac{i\hbar}{2mc\beta + 2(\mathbf{p} - \mathbf{p}.\mathbf{a}_r\mathbf{a}_r).\boldsymbol{\alpha} + \mathbf{k}_r.\boldsymbol{\alpha} - \mathbf{k}_r}, \end{aligned} \tag{6·5}$$

and similarly

$$U_1(\mathbf{p},n;\mathbf{p}-\mathbf{k}_r,n+1_r) = \left(\frac{2\pi \overline{n_r+1}}{c\Omega k_r}\right)^{\frac{1}{2}} \frac{e\hbar}{\mathbf{a}_r.\boldsymbol{\alpha}} \frac{i\hbar}{2mc\beta + 2(\mathbf{p} - \mathbf{p}.\mathbf{a}_r\mathbf{a}_r).\boldsymbol{\alpha} - \mathbf{k}_r.\boldsymbol{\alpha} + k_r}. \tag{6·6}$$

The eigenvalue problem presented by the equation (3·5) then reduces to the solution of the equation

$$\begin{aligned} \sum_r \{U_1(\mathbf{p},n;\mathbf{p}+\mathbf{k}_r,n-1_r)\,\psi(\mathbf{p}+\mathbf{k}_r,n-1_r;\,U) \\ + U_1(\mathbf{p},n;\mathbf{p}-\mathbf{k}_r,n+1_r)\,\psi(\mathbf{p}-\mathbf{k}_r,n+1_r;\,U)\} \\ = U\psi(\mathbf{p},n;\,U) \quad \left(\text{all } \mathbf{p} = \frac{2\pi\hbar}{\Omega^{\frac{1}{3}}}\mathbf{n}\right). \end{aligned} \tag{6·7}$$

The usual iterative method of solution, which is equivalent to the application of ordinary perturbation theory, leads to immediate difficulties with (6·7), because the diagonal elements of U_1^2 diverge in the λ-representation. In fact, no solution of such a kind is possible, on account of the manifold degeneracy, and if no other method were available one might conclude that $\psi(\lambda, U)$ did not exist and the formulae (3·7) were meaningless for that reason, in relation to this example. It is, however, easy to construct an ordered set of solutions in the following way.

Disregarding the trivial solution

$$\psi^{(0)}(\mathbf{p}, n; 0) = \left.\begin{matrix} 0, & n_r \neq 0, \text{ any } r, \\ 1, & n_r = 0, \text{ all } r, \end{matrix}\right\} \tag{6·8}$$

corresponding to the eigenvalue $U = 0$, one is led to consider first

$$\left.\begin{aligned} \psi^{(1)}(\mathbf{p}, n; U) = 0, \quad n_s \neq 0, \quad \text{and } s \neq r, \\ U_1(n_r, n_r-1)\,\psi^{(1)}(n_r-1) + U_1(n_r, n_r+1)\,\psi^{(1)}(n_r+1) = U\psi^{(1)}(n_r), \end{aligned}\right\} \tag{6·9}$$

where
$$\left.\begin{aligned} U_1(n_r, n_r-1) &= \left(\frac{\pi}{2c\Omega k_r}\right)^{\frac{1}{2}} \frac{ie\hbar^2}{\mathbf{a}_r\,.\,\boldsymbol{\alpha}} \frac{n_r^{\frac{1}{2}}}{mc\beta + \mathbf{p}_r\,.\,\boldsymbol{\alpha} - (n_r - \frac{1}{2})\,\mathbf{k}_r\,.\,\boldsymbol{\alpha} - \frac{1}{2}k_r}, \\ U_1(n_r, n_r+1) &= \left(\frac{\pi}{2c\Omega k_r}\right)^{\frac{1}{2}} \frac{ie\hbar^2}{\mathbf{a}_r\,.\,\boldsymbol{\alpha}} \frac{(n_r+1)^{\frac{1}{2}}}{mc\beta + \mathbf{p}_r\,.\,\boldsymbol{\alpha} - (n_r + \frac{1}{2})\,\mathbf{k}_r\,.\,\boldsymbol{\alpha} + \frac{1}{2}k_r}, \end{aligned}\right\} \tag{6·10}$$

and $\mathbf{p}_r$ is the component of $\mathbf{p} + n_r\mathbf{k}_r$ perpendicular to $\mathbf{a}_r$.

The difference equation (6·9), being of the second order, possesses two independent solutions, one of which behaves like $(A_r/U)^{n_r}\,(n_r!)^{-\frac{1}{2}}$, where

$$A_r = \left(\frac{\pi}{2c\Omega}\right)^{\frac{1}{2}} \frac{e\hbar^2}{k_r^{\frac{3}{2}}}, \tag{6·11}$$

and the other like $(U/A_r)^{n_r}\,(n_r!)^{\frac{1}{2}}$, as $n_r \to \infty$. The latter is clearly incompatible with the normalizing condition of (3·5); the former, however, represents a valid solution of (6·7). It has the form

$$\psi^{(1)}(n_r) = \left\{\prod_{m=1}^{n_r} \phi(m)\,U_1(m, m-1)\right\} \psi^{(1)}(0), \tag{6·12}$$

where $\phi(m)$ apparently cannot be written down explicitly, but is defined as a continued matrix fraction

$$\phi(m) = \frac{1}{U - U_1(m, m+1)\,\phi(m+1)\,U_1(m+1, m)}, \tag{6·13}$$

which tends to U^{-1} as $m \to \infty$. This is easily verified by direct substitution of (6·12) into (6·9). The singularity which arises when the denominator in (6·13) has a vanishing eigenvalue, which may happen if U has any of four small values, does not lead to any essential difficulty. For, let $\psi^{(1)}(n_r)$ be defined for $n_r > m$ by

$$\psi^{(1)}(n_r) = \left\{\prod_{l=m+1}^{n_r} \phi(l)\,U_1(l, l-1)\right\} \psi^{(1)}(m);$$

values for $n_r \leqslant m$ can then be derived by substituting successively $n_r = m, m-1$, etc., into (6·9). For example, since, by hypothesis

$$U_1(m, m+1)\,\psi^{(1)}(m+1) = U_1(m, m+1)\,\phi(m+1)\,U_1(m+1, m)\,\psi^{(1)}(m) = U\psi^{(1)}(m)$$

one may infer from (6·9) that $\psi^{(1)}(m-1)$ vanishes, and

$$U_1(m-1,m-2)\,\psi^{(1)}(m-2) = -U_1(m-1,m)\,\psi^{(1)}(m), \quad \text{etc.}$$

The point $U = 0$ is a genuine singularity, and (6·8) is there the only solution. This failure, however, has nothing to do with the divergence difficulty; for the behaviour of the transformation function at $U = 0$ could not make

$$U_1^2(\lambda,\lambda') = \int \psi(\lambda,U)\,U^2\psi(\lambda',U)\,dU$$

diverge without $U_1(\lambda,\lambda')$. The singularity is, in fact, of a weak type which does not affect integration over U.

The behaviour of the function $\psi^{(1)}$, which characterizes processes in which only one kind of photon takes part, has been examined carefully because it is the prototype of $\psi^{(2)}$, $\psi^{(3)}$, $\psi^{(4)}$, etc., which characterize processes in which 2, 3, 4, ... kinds of photons take part. For example, one has for the two-photon processes

$$\psi^{(2)}(\mathbf{p},n;\,U) = 0, \quad n_t \neq 0, \quad \text{any} \quad t \neq r \text{ or } s,$$

$$\begin{aligned} U_1(n_r,n_s;\,n_r-1,n_s)\,\psi^{(2)}(n_r-1,n_s) + U_1(n_r,n_s;\,n_r+1,n_s)\,\psi^{(2)}(n_r+1,n_s) \\ + U_1(n_r,n_s;\,n_r,n_s-1)\,\psi^{(2)}(n_r,n_s-1) + U_1(n_r,n_s;\,n_r\,n_s+1)\,\psi^{(2)}(n_r,n_s+1) \\ = U\psi^{(2)}(n_r,n_s). \end{aligned} \tag{6·14}$$

This equation obviously possesses, among others which do not satisfy the normalizing condition, a solution behaving for large n_r and n_s like $(A_r/U)^{n_r}\,(A_s/U)^{n_s}\,(n_r!\,n_s!)^{-\frac{1}{2}}$, though it is not easy to write down an explicit expression for it; and similarly in the general case there exists a solution which behaves effectively like $\prod_r \left\{\left(\frac{A_r}{U}\right)^{n_r}(n_r!)^{-\frac{1}{2}}\right\}$; to this, according to (3·5), the normalization factor $\exp\left[-\frac{1}{2}\sum_r (A_r/U)^2\right]$ ought to be added, yielding the approximate asymptotic solution

$$\psi(n,\mathbf{p};\,U) = \prod_r \left\{\mathrm{e}^{-\frac{1}{2}(A_r/U)^2}\left(\frac{A_r}{U}\right)^{n_r}(n_r!)^{-\frac{1}{2}}\right\}. \tag{6·15}$$

It is important that the quantity A_r behaves like $k_r^{-\frac{3}{2}}$ for large values of k_r. The formula (6·15) is probably almost good enough for numerical purposes, especially in intense fields, and may be used with confidence in making tests of convergence; it is therefore of great interest to perform the substitution into (3·7). The cross-section, of course, automatically converges; but one obtains in the expression for the self-energy

$$\begin{aligned} \sum_{\lambda'} |\,\psi(\lambda',U)\,|^2\,E_1' &= c\sum_n \prod_r \left\{\mathrm{e}^{-(A_r/U)^2}\left(\frac{A_r}{U}\right)^{2n_r}\frac{1}{n_r!}\right\}\left[\pm\{m^2c^2+(\mathbf{p}_0-\sum_r n_r k_r)^2\}^{\frac{1}{2}}+\sum_r n_r k_r\right] \\ &\sim 2c\sum_r\left(\frac{A_r}{U}\right)^2 k_r, \end{aligned}$$

neglecting $\mathbf{p}_0$ and mc in comparison with n_r and k_r, or

$$\sum_{\lambda'} |\,\psi(\lambda',U)\,|^2\,E_1' \div \frac{2c\Omega}{(2\pi\hbar)^3\,U^2}\int A_r^2 k_r\,dk_r. \tag{6·16}$$

This integral diverges with the upper limit of k_r, and the transverse self-energy in the orthodox theory is therefore not made finite by these refinements. By introducing Dirac's negative energy photons, however, and thereby allowing n to assume negative values, the singularity is reduced to zero: for it is easily confirmed by retracing the argument of this section that it is equally valid for the negative-energy states, and their contribution to the self-energy therefore cancels the left-hand side of (6·16) exactly. Moreover, it is unnecessary to introduce the λ-limiting process to secure this result. Whether it can be regarded as a completely satisfactory resolution of the difficulty of the infinite self-energy awaits a general decision.

7. APPENDIX 1

The theory of the wave field, where any number of derivatives of the wave vector is allowed, is mostly well known, but as no simple formulation of the symmetrization procedure, which plays an important part in the definition of the angular momentum, seems to exist, and for comparison with the theory given in § 4, it is outlined here. Successive derivatives $\dfrac{\partial u}{\partial x^\alpha}, \dfrac{\partial^2 u}{\partial x^\alpha \partial x^\beta}$, etc., of u (which single symbol may stand for several field variables and their conjugates) are represented by u_α, $u_{\alpha\beta}$, etc.

The field equations, which may be derived from the Lagrangian L by a variational procedure, are

$$\frac{\partial L}{\partial u}-\frac{\partial}{\partial x^\alpha}\frac{\partial L}{\partial u_\alpha}+\frac{\partial^2}{\partial x^\alpha \partial x^\beta}\frac{\partial L}{\partial u_{\alpha\beta}}-\ldots=0. \tag{7·1}$$

As a consequence, the unsymmetrized energy-momentum tensor, defined by

$$\begin{aligned} L^{\alpha\beta} = -Lg^{\alpha\beta}+u^\beta\frac{\partial L}{\partial u_\alpha}+\left(u_{\gamma}^{;\beta}\frac{\partial L}{\partial u_{\gamma\alpha}}-u^\beta\frac{\partial}{\partial x^\gamma}\frac{\partial L}{\partial u_{\alpha\gamma}}\right) \\ +\left(u_{\gamma\delta}^{;\beta}\frac{\partial L}{\partial u_{\gamma\delta\alpha}}-u_{\gamma}^{;\beta}\frac{\partial}{\partial x^\delta}\frac{\partial L}{\partial u_{\alpha\gamma\delta}}+u^\beta\frac{\partial^2}{\partial x^\gamma \partial x^\delta}\frac{\partial L}{\partial u_{\gamma\alpha\delta}}\right)+\ldots \end{aligned} \tag{7·2}$$

satisfies

$$\frac{\partial L^{\alpha\beta}}{\partial x^\alpha}=0. \tag{7·3}$$

Under the Lorentz transformation of § 4, one has, if $u \to u+\omega^{\beta}_{\cdot\alpha}S^{\alpha}_{\cdot\beta}u$,

$$L \to L+\omega^{\beta}_{\cdot\alpha}\left\{S^{\alpha}_{\cdot\beta}\left(u\frac{\partial L}{\partial u}+u_\gamma\frac{\partial L}{\partial u_\gamma}+u_{\gamma\delta}\frac{\partial L}{\partial u_{\gamma\delta}}+\ldots\right)-u_\beta\frac{\partial L}{\partial u_\alpha}-\left(u_{\beta\gamma}\frac{\partial L}{\partial u_{\alpha\gamma}}+u_{\gamma\beta}\frac{\partial L}{\partial u_{\gamma\alpha}}\right)-\ldots\right\},$$

so that the invariance of L implies

$$\begin{aligned} S^{\alpha\beta}\left(u\frac{\partial L}{\partial u}+u_\gamma\frac{\partial L}{\partial u_\gamma}+\ldots\right) = \frac{1}{2}\left(u^\beta\frac{\partial L}{\partial u_\alpha}-u^\alpha\frac{\partial L}{\partial u_\beta}\right) \\ +\frac{1}{2}\left(u^{\beta}_{\cdot\gamma}\frac{\partial L}{\partial u_{\alpha\gamma}}+u_{\gamma}^{;\beta}\frac{\partial L}{\partial u_{\gamma\alpha}}-u^{\alpha}_{\cdot\gamma}\frac{\partial L}{\partial u_{\beta\gamma}}-u_{\gamma}^{;\alpha}\frac{\partial L}{\partial u_{\gamma\beta}}\right)+\ldots. \end{aligned} \tag{7·4}$$

By virtue of (7·1), the left-hand side of (7·4) may be written in the form

$$\frac{\partial}{\partial x^\gamma}\left[S^{\alpha\beta}\left\{u\frac{\partial L}{\partial u_\gamma}+\left(u_\delta\frac{\partial L}{\partial u_{\gamma\delta}}-u\frac{\partial}{\partial x^\delta}\frac{\partial L}{\partial u_{\delta\gamma}}\right)+\ldots\right\}\right].$$

Hence, if

$$L^{\alpha\beta,\gamma} = S^{\alpha\beta}\left\{u\frac{\partial L}{\partial u_\gamma} + \left(u_\delta\frac{\partial L}{\partial u_{\gamma\delta}} - u\frac{\partial}{\partial x^\delta}\frac{\partial L}{\partial u_{\delta\gamma}}\right) + \dots\right\} + \frac{1}{2}\left(u^\alpha\frac{\partial L}{\partial u_{\beta\gamma}} - u^\beta\frac{\partial L}{\partial u_{\alpha\gamma}}\right)$$

$$+\frac{1}{2}\left\{u_\delta{}^{\cdot\alpha}\frac{\partial L}{\partial u_{\delta\beta\gamma}} + u^\alpha{}_{\cdot\delta}\frac{\partial L}{\partial u_{\beta\delta\gamma}} - u^\alpha\frac{\partial}{\partial x^\delta}\frac{\partial L}{\partial u_{\beta\gamma\delta}} - u_\delta{}^{\cdot\beta}\frac{\partial L}{\partial u_{\delta\alpha\gamma}} - u^\beta{}_{\cdot\delta}\frac{\partial L}{\partial u_{\alpha\delta\gamma}} + u^\beta\frac{\partial}{\partial x^\delta}\frac{\partial L}{\partial u_{\alpha\gamma\delta}}\right\}$$

$$+\dots, \tag{7·5}$$

one obtains
$$\frac{\partial L^{\alpha\beta,\gamma}}{\partial x^\gamma} = \tfrac{1}{2}(L^{\alpha\beta} - L^{\beta\alpha}), \quad L^{\alpha\beta,\gamma} + L^{\beta\alpha,\gamma} = 0. \tag{7·6}$$

On account of (7·3) and (7·6), the symmetrized energy-momentum tensor

$$T^{\alpha\beta} = \tfrac{1}{2}(L^{\alpha\beta} + L^{\beta\alpha}) + \frac{\partial}{\partial x^\gamma}(L^{\gamma\alpha,\beta} + L^{\gamma\beta,\alpha}) \tag{7·7}$$

satisfies the same equation
$$\frac{\partial T^{\alpha\beta}}{\partial x^\alpha} = 0 \tag{7·8}$$
as $L^{\alpha\beta}$.

8. Appendix 2 (*Added* 15 *October* 1948)

A simple example suffices to illustrate that the perturbation theory described in the text often gives rapidly convergent results when the standard perturbation series diverges. Suppose $H \equiv \begin{bmatrix} 1 & \lambda \\ \lambda & -1 \end{bmatrix} = \sigma_1 + \lambda\sigma_2$, where σ_1 and σ_2 are Pauli spin matrices. Then, according to (2·2), $E_1 = \sigma_1$ and $V = \lambda\sigma_2$. The equation (2·18) gives

$$\frac{U_1}{i\hbar} = \frac{1}{\sigma_1' - \sigma_1}\lambda\sigma_2 = \frac{\lambda}{\sigma_1' - \sigma_1}\frac{1}{\sigma_2} = \frac{1}{\sigma_2}\frac{\lambda}{\sigma_1 + \sigma_1}$$

$$= \tfrac{1}{2}\lambda\sigma_2\sigma_1 = -\tfrac{1}{2}i\lambda\sigma_3, \tag{8·1}$$

whilst (2·20) leads to

$$\begin{aligned} E_1 + E_2 + V_2 &= \exp(\tfrac{1}{2}i\lambda\sigma_3)(\sigma_1 + \lambda\sigma_2)\exp(-\tfrac{1}{2}i\lambda\sigma_3) \\ &= \exp(i\lambda\sigma_3)(\sigma_1 + \lambda\sigma_2) \\ &= (\lambda\sin\lambda + \cos\lambda)\sigma_1 + (\lambda\cos\lambda - \sin\lambda)\sigma_2. \end{aligned} \tag{8·2}$$

If λ is small, this gives

$$E_1 + E_2 + V_2 = (1 + \tfrac{1}{2}\lambda^2 - \tfrac{1}{8}\lambda^4 + \tfrac{1}{144}\lambda^6 - \dots) + (-\tfrac{1}{3}\lambda^3 + \tfrac{1}{30}\lambda^5 - \dots)\sigma_2. \tag{8·3}$$

Since $H^2 = 1 + \lambda^2$, the exact eigenvalues of H are $\pm(1 + \lambda^2)^{\frac{1}{2}}$, and the standard perturbation series, which is a power series in λ, must be

$$E = (1 + \tfrac{1}{2}\lambda^2 - \tfrac{1}{8}\lambda^4 + \tfrac{1}{16}\lambda^6 - \dots). \tag{8·4}$$

Thus at the first approximation the method here employed gives a better result than the fifth approximation of the customary method. Moreover, the standard perturbation series (8·4) diverges for $|\lambda| > 1$, whereas the sequence

$$\lambda_{k+1} = \frac{\lambda_k\cos\lambda_k - \sin\lambda_k}{\lambda_k\sin\lambda_k + \cos\lambda_k} \tag{8·5}$$

converges to zero for (almost) all values of $\lambda = \lambda_1$. This may be seen most easily by writing $\lambda_k = \tan\theta_k$, $-\frac{1}{2}\pi \leqslant \theta_k \leqslant \frac{1}{2}\pi$, when (8·5) reduces to

$$\theta_{k+1} = (\theta_k + n\pi) - \tan\theta_k, \tag{8·6}$$

where n is an integer, positive, negative, or zero. Unless λ satisfies $\sin\lambda = 0$, $\lambda = n\pi$, this sequence converges rapidly to zero (e.g. $\lambda_1 = 1{\cdot}000$, $\lambda_2 = -0{\cdot}2180$, $\lambda_3 = 0{\cdot}0035$, or again $\lambda_1 = 100$, $\lambda_2 = 0{\cdot}0564$, $\lambda_3 = -0{\cdot}0001$). Hence, apart from some exceptional values, the method converges for all values of λ; and if $\lambda < \pi$ it converges monotonically.

The author is greatly indebted to Professor Max Born for much helpful interest and advice in the preparation of this paper.

REFERENCES

Bethe, H. 1947 *Phys. Rev.* **72**, 339.
Born, M. & Green, H. S. 1947 *Proc. Roy. Soc.* A, **191**, 168.
Born, M. & Green, H. S. 1948 *Proc. Roy. Soc.* A, **192**, 166 (App. I).
Born, M. & Rumer, G. 1931 *Z. Phys.* **69**, 141.
Born, M. & Peng, H. W. 1944 *Proc. Roy. Soc. Edinb.* A, **62**, 40, 92, 127.
Chang, T. S. 1946 *Proc. Camb. Phil. Soc.* **42**, 132.
Dirac, P. A. M. 1942 *Proc. Roy. Soc.* A, **180**, 1.
Dirac, P. A. M. 1947*a* Phys. Proc. Cambridge Conference Report.
Dirac, P. A. M. 1947*b* *The principles of quantum mechanics.* Oxford University Press.
Green, H. S. 1949 *Proc. Camb. Phil. Soc.* **45** (in the Press).
Heisenberg, W. 1934 *Z. Phys.* **90**, 209; **92**, 692.
Heisenberg, W. 1943 *Z. Phys.* **120**, 513, 673.
Pauli, W. 1943 *Rev. Mod. Phys.* **15**, 175.
Pauli, W. 1947 Phys. Soc. Cambridge Conference Report.
Peng, H. W. 1946 *Proc. Roy. Soc.* A, **186**, 119.
Podolsky, B. & Schwed, P. 1948 *Rev. Mod. Phys.* **20**, 40.
de Wet, J. 1948 *Proc. Camb. Phil. Soc.* **44**, 546.

The structure of the reaction zone in a flame

By S. F. Boys, *I.C.I. Research Fellow in the University of London*
and J. Corner, *Armament Research Department, Ministry of Supply*

(*Communicated by Sir John Lennard-Jones, F.R.S.—Received* 17 *September* 1948)

The structure of one-dimensional flames is shown to be completely determined by constants such as those of heat conductivity, of diffusion, and of the homogeneous reaction rates. The mathematical problem in the most general case is intractable, but three simple cases are solved fully by a mathematical method of successive approximations. The three cases are those in which diffusion is neglected compared with heat conductivity and in which reaction velocities of the following types are considered: unimolecular, bimolecular, and the quasi-bimolecular form of a unimolecular reaction at low pressures. The method of mathematical approximation is shown to involve errors of the order of only 10 % in some actual cases, an error which is negligible compared with other uncertainties of the problem. In these simple cases it is possible to solve all details of the structure such as the variation of composition and temperature through the flame.

1. INTRODUCTION

The propagation of flames is an extremely important natural phenomenon, yet the theory of the structure of flames has been very little developed. Here we shall give a contribution to the quantitative calculation of the structure by giving mathematical solutions for some simplified models, which however contain all the essential features of the general problem.

Before proceeding with the detailed calculations it appears worth while to show just what are the difficulties in quantitative theoretical predictions of the structure of any flame. The difficulties are twofold, first, even if we were given all necessary data, the mathematical problem when formulated is a particularly intractable one for either analytic or numerical solutions, and secondly, very little of the necessary data is available for any actual flame. The data necessary are the rates of all the reactions occurring in the flame and their dependences on temperature and concentrations, and also the heat conduction and diffusion constants of all components of the mixture. The necessary data will presumably become more readily available as a result of further experimental measurements and theoretical quantum calculations, but considerable work will have to be done on methods of solution of the mathematical problem. The present calculation provides some progress towards this.

There appears to be no difficulty in the general mathematical theory of the structure and mode of propagation of steadily moving plane flames. The well-known process of reactions evolving heat and new particles, and of heat conduction and diffusion provide a complete explanation of flames which move sufficiently slowly that appreciable pressure waves are not set up. It is to this class that the ordinary designation of flames is usually applied and to which we shall be confining our attention in this paper. If we consider a plane flame in a medium which is moved at such a rate that the flame is stationary all concentrations, etc., will become merely functions of the co-ordinate perpendicular to the flame front, which we shall represent by x. If we examine in turn the rate of accumulation of any component or energy in a thin region δx and equate these to zero, since in this steady state there can be no rate of accumulation, we obtain all the differential equations of the structure of the flame.

Let y_1 represent the temperature and $y_2, y_3, \ldots, y_{n+1}$ all the concentrations (as fractions of the total mass) of all the different chemical species. Then the flux of the (r)th component across any unit cross-section will be a linear combination of all the concentration gradients $-\sum_s \lambda_{rs}\frac{dy_s}{dx}$, where λ_{rs} will be given coefficients of heat conduction or diffusion. Hence the rate of accumulation in δx is $\frac{d}{dx}\Sigma\lambda_{rs}\frac{dy_s}{dx}\delta x$. This applies equally whether we are considering the first component, which is the temperature, or one of the concentrations. The amount of the (r)th component passing across a unit cross-section due to the movement of the medium is My_r, where M is the total flow; M must be a constant for the whole system, since there cannot be any accumulation of total mass. Hence the accumulation of the (r)th component in x must be $-M\frac{dy_r}{dx}\delta x$ due to the difference of the flux at each side of the thin layer. It must be noted that y_1 is something of an exception in this case, since the energy flux contains terms dependent on the work done by the pressure in this case, and for complete accuracy in the general case we must use a term $-M\frac{d}{dx}\sum c_s dy_s$ in which the effects of differing specific heats are also allowed for. Lastly, the rates of accumulation due to the generation in the reactions we can write $\mathscr{R}_r\delta x$ where each of these will be some definite function of all the y_s. Hence we have the equations

$$\frac{d}{dx}\sum_s \lambda_{1s}\frac{dy_s}{dx} - M\frac{d}{dx}\sum_s c_s y_s + \mathscr{R}_1 = 0,$$

$$\frac{d}{dx}\sum_s \lambda_{rs}\frac{dy_s}{dx} - M\frac{dy_r}{dx} + \mathscr{R}_r = 0. \qquad r > 1$$

In addition to this we shall know the values of y_r at $x = -\infty$; they will be just the concentrations in the unreacted mixture which is provided in the particular experiment. The values at $x = \infty$ will be calculable from ordinary equilibrium data together with the heat generated in the reactions. Representing these initial and final values as y_r^i and r_r^f the mathematical problem is to find solutions for the y_r as functions of x which satisfy the differential equations and the boundary conditions, together with the value of M which makes it possible for all these conditions to be satisfied simultaneously. This is actually an eigenvalue problem for the value of M, and even when the values of λ_{rs}, C_s and $\mathscr{R}_r$ are all given it is a singularly intractable problem. It might perhaps be stressed that this is the essential nature of the problem and any attempt at solution for simplified models which do not have this essential nature of M being determined by making the solution of one or more differential equations satisfy the boundary conditions cannot be a valid approximation. Since M is the mass flux of the medium in the steady state it is the velocity of the flame expressed by giving the mass of the medium through which unit area of the flame moves per unit time.

There appear to have been only two previous attempts to make predictions of flame velocities on the basis of models of the flame structure from such considerations. An attempt by Zeldowitsch & Frank-Kamenetzki (1938) is on these lines, reducing the system to one differential equation and then making an estimate of a solution by a temporary neglect of one of the terms of the differential equation. It is hard to estimate the significance of this, but it does not give a systematic method of successive approximation to the solution of the differential equation. In the other attempt, by Lewis & von Elbe (1934), a number of *ad hoc* approximations are introduced into the problem and so far as the authors can trace the proper method of determining M by making the solution fit all the boundary conditions is not used.

The simplified models which we shall treat systematically in this paper are those in which a single reaction occurs, for which we shall consider both unimolecular reactions and bimolecular reactions, and in which diffusion is neglected and only heat conduction considered. The actual circumstances to which such a picture might apply accurately are probably only those of a reaction propagated through a liquid. If an exothermic reaction takes place in the upper layers of a liquid the reaction could be propagated downwards through a metastable liquid in a manner exactly equivalent theoretically to a flame. It is conceivable that such a method might actually be used for the measurement of a particular fast liquid reaction velocity by using one of the formulae of this paper in a reverse sense. However, this was not the aim in performing these calculations. The simple cases were taken because they were possible to solve and would provide formulae for the propagation velocities which were of the correct nature. It was hoped that these could be used as crude approximations for gaseous reactions even though diffusion does occur in these. Actually these formulae did not prove as useful for making predictive estimates for the particular flames in which the authors were interested as had been hoped. Detailed examination indicated that the circumstances were too complicated to be represented by a single reaction. However, the fact that it was possible to carry out the complete analysis for these simple cases appeared to be very helpful in the general understanding of the actual processes in flame, and in estimating what factors were of importance in determining variations of propagation velocities.

We shall now proceed to carry out the mathematical analysis for these simplified models.

2. MATHEMATICAL THEORY OF THE PROPAGATION OF A FLAME BY THERMAL REACTION

We shall examine the steady propagation of a plane reaction zone through a homogeneous medium with the following properties:

(i) the medium is capable of a single reaction whose rate at any point in the medium is determined solely by the temperature, pressure, and the stage to which the reaction has already proceeded at the point;

(ii) diffusion is small enough to be neglected in the time concerned;

(iii) the velocity of movement of the reaction zone is sufficiently slow that the pressure is effectively constant through this zone.

The general analysis would apply to a flame in a gas, or the movement of a reaction zone in a solid or liquid provided the above conditions were fulfilled. We shall examine the general case, and then some special cases in which constant specific heats and thermal conductivities, together with activation energy formulae for the reaction rates, are assumed. The general case can always be integrated numerically when the necessary data are available, and in the special cases explicit formulae for the rate of burning can be obtained.

We examine the structure of a steadily moving plane reaction zone. We take an x axis of reference perpendicular to the plane, so that the condition of the medium is dependent only on x, and independent of y and z. We take an origin moving with the reaction zone, so that relative to our axes of reference the unreacted medium moves from the direction of negative x, and after passing through the reaction zone the products pass towards $x = \infty$. All properties of the system are functions of x. When U is the velocity of the medium at any point, the rate at which any property X of a small given portion of the medium varies with time is given by

$$\frac{dX}{dt} = U\frac{dX}{dx}. \tag{1}$$

U will alter as we pass through the reaction zone.

The quantities whose variation through the reaction zone as required are U the velocity, V the volume per unit mass, T the temperature, and ϵ the fraction the reaction has progressed towards completion. We can immediately obtain a relation between U and V. If we consider the material moving along a cylinder with walls parallel to the x axis and of unit cross-section, it is apparent that the same mass is moving across any cross-section of this per unit time, and representing this by M, we have

$$M = \frac{U}{V}. \tag{2}$$

Since the pressure is constant, V must be a function of T and ϵ only, known from the equation of state for the medium with a given ϵ.

The energy crossing any normal section of the cylinder must be a constant, or energy would accumulate or disappear between two cross-sections. The energy flow is composed of:

(*a*) the intrinsic energy E per unit mass, transported by the mass flow;

(*b*) the work done by the material on one side of the section by its pressure on the material on the other side of the section;

(*c*) the flow of heat due to thermal conductivity.

The flow by conduction at large distances from the reaction zone will be zero on both sides. Denote the conditions at a large distance before the reaction zone by T_0, V_0, etc., and at large distances past as T_m, V_m, etc. Then we have

$$ME + PU - \frac{\lambda dT}{dx} = ME_m + PU_m = ME_0 + PU_0, \tag{3}$$

which gives $$\frac{\lambda}{M}\frac{dT}{dx} = E - E_m + P(V - V_m) = H - H_m \quad \text{or} \quad H - H_0, \tag{4}$$

H is the total heat content per unit mass.

We need a further relation, which is supplied by consideration of reaction velocity. The homogeneous rate, expressed as $d\epsilon/dt$, will be a function of ϵ, V and T, by our initial assumptions. Hence we have

$$U\frac{d\epsilon}{dx} = \frac{d\epsilon}{dt} = \mathscr{R}(\epsilon, V, T). \tag{5}$$

Equations (4) and (5) can be written in terms of ϵ and T alone, since H will be a known function of ϵ, T and V, and the latter can be eliminated by using the equation of state. In this way we can obtain two differential equations in ϵ, x and T, which can be simplified by dividing one by the other, giving

$$\frac{d\epsilon}{dT} = \frac{\lambda\mathscr{R}(\epsilon, T)}{M^2V(H - H_m)}. \tag{6}$$

This is possible only because x does not appear except as dx. In more general cases than we are considering in this paper, x may appear explicitly; this happens when, for example, the phenomena in the flame zone depend on the distance from some solid boundary. In such cases there is no point in forming (6). In our case, however, this step is useful because the problem can be solved from equation (6) alone. If distances in the flame zone are of interest the results can be substituted in (4), giving the relation between x and T or ϵ.

We require the integral of (6) which is such that $T = T_0$ when $\epsilon = 0$, and $T = T_m$ when $\epsilon = 1$. T_0 is the temperature at a large distance inside the unreacted medium, the initial temperature, which is known in any particular set of experiments. T_m follows from this by the relation

$$H_m - H_0 = 0,$$

where H_m is the total heat content for $\epsilon = 1$, and H_0 refers to $\epsilon = 0$.

T_m is the temperature of burning at constant pressure, easily calculated from thermochemical data.

Equation (6) is a first-order equation, and its solution is determined by the parameters of the equation, together with one pair of corresponding values of ϵ and T. The solution determined by the condition at T_0 will in general not fulfil the condition at the upper limit T_m. There will be one value of M for which both conditions are satisfied; this determines the only velocity possible for a steadily moving reaction zone. This value of M is a function of the other parameters in the equation.

The form of the equations ensures that ϵ and T remain constant beyond the point where $\epsilon = 1$ and $T = T_m$. This has the result that we can restrict our attention to the equation for $d\epsilon/dT$, when trying to find the correct value of M. No matter whether the point $\epsilon = 1$, $T = T_m$ is at a finite or an infinite distance, we can be sure that everywhere beyond this point the outgoing gases will be at the temperature T_m and will be completely reacted, which is necessary for a physically satisfactory reaction zone. At distances which are large compared with the effective thickness of the flame, the gas may be subjected to cooling and consequent changes of composition, but these cannot affect the rate of burning. The only influences which can have any effect on the rate of burning are those which can act in the region where by far the greater part of the reaction takes place.

As the problem is solved by choosing an arbitrary parameter in a differential equation so that it has a solution which satisfies several boundary conditions, the problem is essentially the determination of an eigenvalue. The very general form in which we have left the various functions such as the heat content data, makes no difference to the principles of solution of the problem. We shall of course make much more special assumptions in following sections, but we shall not need to introduce any other general ideas of the reaction zone.

The velocity of the reaction relative to the unburnt medium is MV_0. The corresponding relation between ϵ and T is the solution of (6) with the proper value of M; finally, the (x, T) relation can be found by substituting in (4). This provides a scale of distances for the variation of T and ϵ through the reaction zone.

3. A SOLUTION BY SUCCESSIVE APPROXIMATIONS

We shall now examine three simple cases which can be integrated explicitly by a method of successive approximations. These cases correspond to simple forms of the reaction rate which are frequently found in practice. The method of successive approximations is not necessarily restricted to these cases.

We shall limit our examination to a zone in perfect gases with constant specific heats per unit mass. All these assumptions will be discussed fully in appendices.

If Q_m is the heat of reaction at constant pressure at the temperature T_m of the final products, we have

$$H-H_m = (1-\epsilon)\,Q_m + (T-T_m)\,[C_p^1(1-\epsilon)+\epsilon C_p], \tag{7}$$

where C_p^1 is the specific heat of the reactant and C_p that of the products, both at constant pressure. It would be possible to integrate this case if the problem required this refinement, but since a simpler case is adequate for our purposes we shall put $C_p^1 = C_p = C$. We can also write Q instead of Q_m, since now the heat of reaction is independent of the temperature. Thus finally

$$H-H_m = (1-\epsilon)\,Q + (T-T_m)\,C, \tag{8}$$

so that

$$\frac{\lambda}{M}\frac{dT}{dx} = Q(1-\epsilon) + C(T-T_m), \tag{9}$$

and substituting in (6),

$$\frac{d\epsilon}{dT} = \frac{\lambda\mathscr{R}(\epsilon, T)}{M^2V\{(1-\epsilon)\,Q+(T-T_m)\,C\}}. \tag{10}$$

In a simple reaction in which molecules of molecular weight W give molecules of molecular weight w, the equation of state is

$$PV = \left\{\frac{1-\epsilon}{W}+\frac{\epsilon}{w}\right\}RT = (1+n\epsilon)\frac{RT}{W}, \tag{11}$$

where $n = \dfrac{W}{w} - 1$, and R is the gas constant per mole.

If the reactant consists of more than one type of molecule with different molecular weights, this equation will have the same form but with a different value of n.

We shall derive formulae for the rate of burning associated with three types of homogeneous reaction: (I) a first-order rate from a unimolecular mechanism; (II) a second-order rate from a bimolecular reaction; (III) a second-order rate as the low pressure result of a unimolecular breakdown. These are described by the values of $d\epsilon/dt$ as functions of ϵ, V and T.

$$\text{(I)} \quad \mathscr{R}(\epsilon, V, T) = \frac{d\epsilon}{dt} = B_1(1-\epsilon)\exp(-A/RT). \tag{12}$$

$$\text{(II)} \quad \mathscr{R}(\epsilon, V, T) = \frac{d\epsilon}{dt} = B_2(1-\epsilon)^2 V^{-1}\exp(-A/RT). \tag{13}$$

$$\text{(III)} \quad \mathscr{R}(\epsilon, V, T) = \frac{d\epsilon}{dt} = B_3(1-\epsilon)(1+n\epsilon) V^{-1}\exp(-A/RT). \tag{14}$$

The values of the constants will be discussed when the results are applied to particular cases.

Substituting in (10), we have

$$\text{(I)} \quad \frac{d\epsilon}{dT} = \frac{\lambda B_1(1-\epsilon)\exp(-A/RT)}{M^2 V\{Q(1-\epsilon)+C(T-T_m)\}}. \tag{15}$$

$$\text{(II)} \quad \frac{d\epsilon}{dT} = \frac{\lambda B_2(1-\epsilon)^2\exp(-A/RT)}{M^2 V^2\{Q(1-\epsilon)+C(T-T_m)\}}. \tag{16}$$

$$\text{(III)} \quad \frac{d\epsilon}{dT} = \frac{\lambda B_3(1-\epsilon)(1+n\epsilon)\exp(-A/RT)}{M^2 V^2\{Q(1-\epsilon)+C(T-T_m)\}}. \tag{17}$$

The essentials of the method of integration are the same in all cases, but case (III) is simplest in algebraic detail, and we shall consider this first. From (11) and (17), the relation between ϵ and T is

$$\frac{d\epsilon}{dT} = \frac{\lambda B_3 P^2 W^2(1-\epsilon)\exp(-A/RT)}{M^2 R^2 T^2(1+n\epsilon)\{Q(1-\epsilon)+C(T-T_m)\}}. \tag{18}$$

This equation cannot be integrated exactly. Near the outer boundary of the reaction zone $T-T_m$ and $1-\epsilon$ are small, and these terms determine the behaviour of the solution in this region. All other functions of T and ϵ can be given the values corresponding to $T = T_m$ and $\epsilon = 1$. The equation thus simplified has a solution which at large distances, where $T-T_m$ and $1-\epsilon$ are sufficiently small, approaches that solution of the original equation satisfying the boundary condition that $T = T_m$ when $\epsilon = 1$. This (ϵ, T) relation will then be used in the awkward coupling term to make (18) integrable by quadratures. This provides us with a second approximation to the solution of (18). We shall not go beyond this order of approximation, because comparison with step-by-step integration of the differential equation has shown that the accuracy is ample for our purpose; moreover, the next approximation equation would be as difficult to solve analytically as the original equation itself. It is not possible to stop at the extremely simple first approximation, because the error in the rate of burning, as compared with an accurate numerical solution of the equations, is a factor of about three. The error in the second approximation has not been more than 15 % in those examples which we have tested; these belonged to case (II), and the errors are presumably of the same order of magnitude in cases (I) and (III).

Write $$1-\epsilon = \xi \quad \text{and} \quad T_m - T = \eta.$$

The first approximation is the solution of the equation

$$\frac{d\xi}{d\eta} = \frac{\lambda B_3 P^2 W^2 \xi \exp(-A/RT_m)}{M^2 R^2 T_m^2 (1+n)(Q\xi - C\eta)} = \frac{D_3 \xi}{Q\xi - C\eta}, \tag{19}$$

where D_3 is defined by this equation. Hence

$$\frac{d\eta}{d\xi} + \frac{C}{D_3}\frac{\eta}{\xi} = \frac{Q}{D_3}. \tag{20}$$

The boundary condition $T = T_m$ when $\epsilon = 1$, is $\eta = 0$ when $\xi = 0$. The solution of (20) which satisfies this condition is

$$\eta = \frac{Q\xi}{C+D_3}. \tag{21}$$

Inserting this in the coupling term of (18), we have as the equation for the second approximation,

$$\frac{d\epsilon}{dT} = \frac{\lambda B_3 P^2 W^2 (C+D_3)}{M^2 R^2 (1+n\epsilon) Q D_3}\frac{\exp(-A/RT)}{T^2},$$

of which the integration is immediate:

$$[\epsilon + \tfrac{1}{2}n\epsilon^2]_1^{\epsilon} = \frac{\lambda B_3 P^2 W^2 (C+D_3)}{M^2 R^2 Q D_3}\frac{R}{A}[\exp(-A/RT)]_{T_m}^{T}. \tag{22}$$

This contains one parameter which is still arbitrary, namely M. This we must choose so that the solution satisfies the boundary condition at the beginning of the flame, that $\epsilon = 0$ when $T = T_0$. When T is small, e^{-A/RT_0} is effectively zero in comparison with e^{-A/RT_m}, and so the equation which must be satisfied by M is

$$1 + \frac{n}{2} = \frac{\lambda B_3 P^2 W^2 (C+D_3)}{M^2 R^2 Q D_3}\left(\frac{R}{A}\right)\exp\left\{-\frac{A}{RT_m}\right\},$$

and using (19), the definition of D_3, we find

$$M^2 = \frac{\lambda B_3 P^2 W^2 \exp(-A/RT_m)}{R^2 T_m^2 C(1+n)\left\{\left(\frac{1+\frac{1}{2}n}{1+n}\right)\left(\frac{A}{RT_m}\right)\left(\frac{Q}{CT_m}\right)-1\right\}}. \tag{23}$$

This gives M, the mass of the medium which passes through unit cross-section of the reaction zone per unit time. V_0 is the specific volume of the original reactant, and MV_0 is the velocity of movement of the reaction zone relative to the unreacted medium. To obtain the relation between ϵ and T through the flame, the value of M from (23) would be used in the solution (22). By solving a quadratic the solution can be written as an explicit function $\epsilon = f(T)$.

In using this equation and similar ones which we shall obtain later, one point must be noted. The gas constant R appears from two sources; as it enters from (11), the equation of state of the gas, it has to be in the unit

$$\frac{(\text{unit of pressure})\times(\text{unit of volume})}{\text{unit of temperature}}\text{ of a mole of gas.}$$

In other words, R is in the appropriate mechanical units. The activation energy A is conventionally given in heat units, as calories, so the R associated with the A is in heat units, calories per mole per degree. To prevent possible confusion, we have written all the formulae in terms of R and A/R, which has the dimensions of temperature.

Case (I) can be treated similarly, and if we make a further small approximation the answer can be written in terms of functions which are as simple as in the case we have just discussed. The accurate relation between ϵ and T is

$$\frac{d\epsilon}{dT}=\frac{\lambda B_1PW(1-\epsilon)\exp(-A/RT)}{M^2RT(1+n\epsilon)\{Q(1-\epsilon)+C(T-T_m)\}}, \tag{24}$$

derived from (15) and (11).

For $T-T_m$ and $1-\epsilon$ small the equation takes the form

$$\frac{d\xi}{d\eta}=\frac{D_1\xi}{Q\xi-C\eta}, \tag{25}$$

where
$$D_1=\frac{\lambda B_1PW\exp(-A/RT_m)}{M^2RT_m(1+n)}, \tag{26}$$

giving
$$T_m-T=\frac{Q}{C+D_1}(1-\epsilon), \tag{27}$$

and the equation for the second approximation is

$$\frac{d\epsilon}{dT}=\frac{\lambda B_1PWT}{M^2R(1+n\epsilon)}\left(\frac{C+D_1}{QD_1}\right)\left\{\frac{\exp(-A/RT)}{T^2}\right\}. \tag{28}$$

This can be integrated exactly, the answer involving the exponential integral, which is a tabulated function. An approximate solution, from which the behaviour is more obvious, can be obtained by replacing T in the numerator by T_m. This approximation cannot be a serious source of error, since in any case the representation of the reaction rate as $B\,e^{-A/RT}$ is itself an approximation which may not be very close at the temperatures involved. Moreover, the exponential integral of the accurate solution can be expanded in an asymptotic series which shows that the error is of the order of a few per cent for the values of A/RT_m which we shall consider.

Making this approximation, then, the integration proceeds as before, and the equation for the rate of burning is

$$M^2=\frac{\lambda B_1PW\exp(-A/RT_m)}{CRT_m(1+n)\left\{\left(\frac{1+\frac{1}{2}n}{1+n}\right)\left(\frac{A}{RT_m}\right)\left(\frac{Q}{CT_m}\right)-1\right\}}. \tag{29}$$

The presence of a factor $(1-\epsilon)^2$ in the case (II) equation makes its treatment more complicated. The method is the same as before. The exact equation is

$$\frac{d\epsilon}{dT} = \frac{\lambda B_2 P^2 W^2 (1-\epsilon)^2 \exp(-A/RT)}{M^2 R^2 T^2 (1+n\epsilon)^2 \{Q(1-\epsilon) + C(T-T_m)\}}, \tag{30}$$

and the equation for the first approximation is

$$\frac{d\xi}{d\eta} = \frac{D_2 \xi^2}{Q\xi - C\eta}, \tag{31}$$

where $$D_2 = \frac{\lambda B_2 P^2 W^2 \exp(-A/RT_m)}{M^2 R^2 T_m^2 (1+n)^2}.$$

The equation can be written as

$$\frac{d\eta}{d\xi} + \frac{C\eta}{D_2 \xi^2} = \frac{Q}{D_2 \xi}, \tag{32}$$

of which the solution passing through $\xi = \eta = 0$ is

$$\eta \exp\left(-\frac{C}{D_2 \xi}\right) = \frac{Q}{D_2}\int_0^{\xi} \exp\left(-\frac{C}{D_2 \xi}\right)\frac{d\xi}{\xi}. \tag{33}$$

The exponential integral is defined (e.g. Jahnke & Emde, 1933, p. 1) as

$$Ei(-x) = \int_{\infty}^{x} \mathrm{e}^{-t}\, dt/t. \tag{34}$$

Let $\psi = C/D_2$. Transforming the variable of integration in (33) to ψ/ξ we see that the solution (33) can be written as

$$T_m - T = -\frac{Q}{D_2}\exp\left(\frac{\psi}{\xi}\right) Ei\left(-\frac{\psi}{\xi}\right). \tag{35}$$

The exponential integral has been tabulated in several places.

Substituting in the $T - T_m$ term of (30),

$$\frac{d\epsilon}{dT} = \frac{\lambda B_2 P^2 W^2}{M^2 R^2 Q (1+n\epsilon)^2} \frac{\exp(-A/RT)}{T^2} \frac{\xi^2}{\{\xi + \psi \exp(\psi/\xi)\, Ei(-\psi/\xi)\}}.$$

Hence $$\int_{1-\epsilon}^{0} \{\xi + \psi \exp(\psi/\xi)\, Ei(-\psi/\xi)\}\{(1+n)/\xi - n\}^2 d\xi$$
$$= \frac{\lambda B_2 P^2 W^2}{M^2 R^2 Q}\left(\frac{R}{A}\right)[\exp(-A/RT) - \exp(-A/RT_m)]. \tag{36}$$

This solution satisfies the boundary condition at T_m, but not necessarily at the beginning of the flame. Imposing the condition that $\epsilon = 0$ when $T = T_0$, gives as the equation to determine M,

$$\int_0^1 (\xi + \psi \exp(\psi/\xi)\, Ei(-\psi/\xi)\}\{(1+n)/\xi - n\}^2 d\xi = \frac{\lambda B_2 P^2 W^2}{M^2 R^2 Q}\left(\frac{R}{A}\right)\exp(-A/RT_m).$$

Write $\theta = \psi/\xi$. The result is

$$(1+n)^2 g_1(\psi) - 2n(1+n)\, g_2(\psi) + n^2 g_3(\psi) = \frac{\psi \lambda B_2 P^2 W^2}{M^2 R^2 Q}\left(\frac{R}{A}\right)\exp(-A/RT_m), \tag{37}$$

where
$$\left.\begin{aligned} g_1(\psi) &= \psi\int_\psi^\infty \{1/\theta + e^\theta Ei(-\theta)\}\,d\theta, \\ g_2(\psi) &= \psi^2\int_\psi^\infty \{1/\theta + e^\theta Ei(-\theta)\}\,d\theta/\theta, \\ g_3(\psi) &= \psi^3\int_\psi^\infty \{1/\theta + e^\theta Ei(-\theta)\}\,d\theta/\theta^2. \end{aligned}\right\} \tag{38}$$

Simplifying the right of (37) by using the definitions of ψ and D_2, it becomes just $(CT_m/Q)(RT_m/A)(1+n)^2$, and (37) becomes

$$g_1(\psi) - \frac{2n}{1+n}g_2(\psi) + \left(\frac{n}{1+n}\right)^2 g_3(\psi) = (CT_m/Q)(RT_m/A). \tag{39}$$

We have tabulated the functions g_1, g_2 and g_3 for the whole range of ψ likely to be used. These are given in table 1.

(39) is an equation to determine ψ, and when its value for a particular case has been determined the value of M is found from

$$\psi = C/D_2,$$

giving
$$M^2 = \lambda B_2 P^2 W^2 \psi \exp(-A/RT_m)/CR^2T_m^2(1+n)^2. \tag{40}$$

It should be noted that (39) does not contain P, so for a given reaction ψ is independent of the pressure, and M proportional to the pressure.

The accuracy of the second approximation can be tested by comparison with the results of numerical integration. This has been done only for case (II), but there is no reason to suppose that the results would be any less accurate in other cases; the physical justification for the method of approximation is the same for all. The error seems to depend chiefly on A/RT_m; for values near 8, the rate of burning was found to be too low by about 6 %, this varying by 2 to 3 % from case to case. For A/RT_m near 5, the rate was too low by an error of 10 to 15 %.

For cases (I) and (III) there is an obvious relation between ϵ and T in the second approximation. In case (II) the result is

$$\begin{aligned}(1-\epsilon)g_1(\psi/(1-\epsilon)) - \frac{2n(1-\epsilon)^2}{1+n}g_2\left(\frac{\psi}{1-\epsilon}\right) + \frac{n^2(1-\epsilon)^3}{(1+n)^2}g_3\left(\frac{\psi}{1-\epsilon}\right) \\ = \left(\frac{CT_m}{Q}\right)\left(\frac{RT_m}{A}\right)\left[1 - \exp\left(\frac{A}{RT_m} - \frac{A}{RT}\right)\right].\end{aligned} \tag{41}$$

For any set of values of A, T_m, Q, C, n, equation (39) gives ψ, which can be used in (41) to find the T corresponding to any given ϵ. The error in the (ϵ, T) curve is small; comparison with a numerical integration carried out with the accurate M showed that ϵ was too large by 3 % at $\epsilon = 0{\cdot}5$, and the error was everywhere of the same amount except at low temperatures.

This (ϵ, T) relation can be used to find a value of M which is several per cent better than that of the second approximation. For this we choose $\epsilon_1, \epsilon_2 \simeq 0{\cdot}4, 0{\cdot}6$ respectively, and such that $\psi/(1-\epsilon_1)$ and $\psi/(1-\epsilon_2)$ occur in table 1. (41) gives the corresponding T_1, T_2 and hence $d\epsilon/dT$ at $\frac{1}{2}(T_1+T_2)$. Substitution in the accurate equation of the zone gives M.

TABLE 1. FUNCTIONS FOR THE SOLUTION OF THE CASE (II) EQUATION

ψ	g_1	g_2	g_3
4·5	0·840	0·397	0·258
3·9	0·822	0·386	0·250
3·3	0·800	0·372	0·240
2·7	0·771	0·355	0·228
2·1	0·731	0·332	0·211
1·5	0·672	0·298	0·188
1·4	0·660	0·291	0·183
1·3	0·646	0·283	0·178
1·2	0·631	0·274	0·172
1·1	0·615	0·265	0·165
1·0	0·596	0·255	0·159
0·95	0·586	0·249	0·155
0·9	0·576	0·244	0·151
0·85	0·565	0·238	0·147
0·8	0·553	0·231	0·143
0·75	0·541	0·224	0·138
0·7	0·527	0·217	0·133
0·65	0·512	0·210	0·128
0·6	0·497	0·202	0·123
0·55	0·480	0·193	0·117
0·5	0·461	0·183	0·111
0·45	0·442	0·173	0·104
0·4	0·419	0·162	0·097
0·38	0·410	0·158	0·094
0·36	0·400	0·153	0·091
0·34	0·389	0·148	0·088
0·32	0·378	0·142	0·085
0·30	0·367	0·137	0·081
0·28	0·355	0·131	0·077
0·26	0·342	0·125	0·074
0·24	0·328	0·119	0·070
0·22	0·314	0·112	0·066
0·20	0·299	0·106	0·061
0·18	0·282	0·098	0·057
0·16	0·265	0·091	0·052
0·14	0·245	0·082	0·047
0·12	0·224	0·074	0·042
0·10	0·201	0·064	0·036
0·08	0·176	0·054	0·030
0·06	0·146	0·043	0·024

When the (ϵ, T) relation has been found, it can be used in the equation for the heat-conduction (9), to find the values of x corresponding to various T. We have imposed the condition, necessary for a physically possible reaction zone, that $T = T_m$ and $\epsilon = 1$ together. It is immaterial whether this point has a finite or an infinite value of x, though usually the latter is the case. The flame is then theoretically infinite, though most of the reaction takes place within a very small distance of the surface. If $\epsilon = 1$ and $T = T_m$ at a finite value of x, the form of the equations ensures that ϵ and T remain constant for all larger values of x. At distances from the surface which are large compared with the effective thickness of the flame, the gas may be subjected to cooling and subsequent change of composition, but these cannot affect the rate of burning, no matter whether the flame thickness is finite or not. The only influences which can have any effect on the rate of burning are those which can act in the region where all but a small fraction of the chemical change takes place.

First approximations to the (x, T) relation can be found from the first approximation to the (ϵ, T) law.

Cases (*I*) *and* (*III*). The first approximation is (21) or (27), and using this in (9)

$$\frac{\lambda}{M}\frac{dT}{dx} = D(T_m - T),$$

with $D = D_1$ or D_3 in the two cases.

Hence $T_m - T$ and $1 - \epsilon$ behave at large distances like multiples of $\exp(-MDx/\lambda)$. The thickness of the reaction zone is, strictly speaking, infinite.

Case (*II*). From (31) and (9),

$$d\xi/dx = -MD_2\xi^2/\lambda,$$

with the solution $\quad (1-\epsilon)^{-1} = \xi^{-1} = MD_2x/\lambda + \text{constant},$

and an infinite flame thickness.

To get a second approximation in these cases, it is easiest to tabulate ϵ as a function of T, given by the appropriate second approximation, and then carry out a numerical integration of

$$x = \int_{T_1}^{T} \frac{\lambda dT}{M\{Q(1-\epsilon) + C(T - T_m)\}},$$

where T_1 is the temperature at the arbitrarily chosen origin of x.

In all these cases the 'thickness of the flame' is in the first approximation proportional to λ/MD.

In case (II), $$\frac{\lambda}{MD_2} = RT_m(1+n)\,[\lambda\psi/CB_2 \exp(-A/RT_m)]^{\frac{1}{2}}/PW. \tag{42}$$

Using (23) and (19) for case (III),

$$\lambda/MD_3 = (RT_m/PW)\left[\lambda(1+n)/CB_3 \exp(-A/RT_m)\left\{\left(\frac{1+\frac{1}{2}n}{1+n}\right)\left(\frac{A}{RT_m}\right)\left(\frac{Q}{CT_m}\right) - 1\right\}\right]^{\frac{1}{2}}.$$

In case (I),

$$\lambda/MD_1 = \left[\lambda RT_m(1+n)/B_1PW \exp(-A/RT_m)\left\{\left(\frac{1+\frac{1}{2}n}{1+n}\right)\left(\frac{A}{RT_m}\right)\left(\frac{Q}{RT_m}\right) - 1\right\}\right]^{\frac{1}{2}}.$$

4. SEPARATION OF CERTAIN PARAMETERS

It is possible to obtain explicit formulae for the effect of certain of the parameters of the problem by a method which is simple and exact. We take the equations of case (I) as an example:

$$\frac{d\epsilon}{dx} = B_1(1-\epsilon)\,PW \exp(-A/RT)/MRT(1+n\epsilon), \tag{43}$$

$$\frac{dT}{dx} = M\{Q(1-\epsilon) + C(T - T_m)\}/\lambda, \tag{44}$$

which can be written as

$$(\lambda/B_1PW)^{\frac{1}{2}}\frac{d\epsilon}{dx} = (\lambda B_1PW/M^2)^{\frac{1}{2}}f(\epsilon, T, A, n), \tag{45}$$

$$(\lambda/B_1PW)^{\frac{1}{2}}\frac{dT}{dx} = (\lambda B_1PW/M^2)^{-\frac{1}{2}}g(\epsilon, T, A, n), \tag{46}$$

where f and g are functions of their arguments which are obvious from (43) and (44), and whose precise form will not be needed hereafter. This pair of equations connects ϵ and T with $x(B_1PW/\lambda)^{\frac{1}{2}}$ and the only parameter left in them is $(M^2/\lambda B_1PW)^{\frac{1}{2}}$. The boundary conditions fix the values of ϵ and T at two points, where $x(B_1PW/\lambda)^{\frac{1}{2}}$ has fixed values (0 and ∞).

The values of ϵ and T at these points are independent of B_1, P, W, and λ; it follows that the solution of (45) and (46) which satisfies such boundary conditions has

$$\frac{M}{(\lambda B_1PW)^{\frac{1}{2}}} = \text{function of } (A, n, Q, C, T_m), \tag{47}$$

and any particular value of T or ϵ occurs at a value of $x(B_1PW/\lambda)^{\frac{1}{2}}$ which depends only on A, n, Q, C, T_m.

In other words, the rate of burning $M = (\lambda B_1PW)^{\frac{1}{2}} \times$ function of (A, n, Q, C, T_m), and any dimension (of the reaction zone) which is defined by values of ϵ and T, will be proportional to

$$\left(\frac{\lambda}{B_1PW}\right)^{\frac{1}{2}} \times \text{function of } (A, n, Q, C, T_m). \tag{48}$$

The thickness of the complete zone of burning is infinite, but it is convenient to speak of a 'flame thickness', which in this case is proportional to the quantity (48). In particular, the pressure variation of the rate of burning is

$$M \propto P^{\frac{1}{2}}. \tag{49}$$

The same process can be applied to cases (II) and (III), and the results are

$$M = PW(\lambda B)^{\frac{1}{2}} \times \text{function of } (A, n, Q, C, T_m), \tag{50}$$

$$\text{'flame thickness'} = \frac{1}{PW}(\lambda/B)^{\frac{1}{2}} \times \text{function of } (A, n, Q, C, T_m), \tag{51}$$

with $B = B_2$ or B_3.

It will be noticed that the rate of burning is proportional to $\lambda^{\frac{1}{2}}$ for all three chemical mechanisms. This suggests that the result may be true under more general assumptions. We shall now show that our fundamental equations, without any assumption other than that the conductivity is constant, lead to the results that the rate of burning and the 'flame thickness' are both proportional to $\lambda^{\frac{1}{2}}$.

Equation (4) is
$$\frac{\lambda}{M}\frac{dT}{dx} = H - H_m, \tag{52}$$

and (5) is
$$VM\frac{d}{dx} = \mathcal{R}(\epsilon, V, T), \tag{53}$$

which can be written as

$$\lambda^{\frac{1}{2}}\frac{dT}{dx} = \frac{M}{\lambda^{\frac{1}{2}}}(H - H_m), \quad \lambda^{\frac{1}{2}}\frac{d\epsilon}{dx} = \frac{\lambda^{\frac{1}{2}}}{M}\frac{\mathscr{R}(\epsilon, V, T)}{V},$$

and it follows that the rate of burning and flame thickness are proportional to $\lambda^{\frac{1}{2}}$.

5. The method of numerical integration

In § 3 we gave approximate solutions for the equations of the reaction zone. These solutions are easily handled, but the method would fail if the equations were substantially more complicated, as, for example, if the thermal conductivity were a function of the temperature. In such a case it would be necessary to solve the equations numerically.

The reaction zone equation is very easily integrated numerically. Consider, for example, case (II), whose equation is (30) of § 2:

$$\frac{d\epsilon}{dT} = \frac{\lambda B_2 P^2 W^2 (1-\epsilon)^2 \exp(-A/RT)}{M^2 R^2 T^2 (1+n\epsilon)^2 \{Q(1-\epsilon) + C(T - T_m)\}}.$$

This can be integrated by a step-by-step process, for any assumed value of M. Starting from one of the boundary conditions, the solution can be carried sufficiently near to the other extreme temperature of the reaction zone, to see whether the solution satisfies the other boundary condition. For example, suppose that the integration has been started at the cool side of the flame, where ϵ is practically zero and $T = T_0$. If M is less than a certain critical value, dT/dx becomes negative after a certain value of T has been reached in the integration. $d\epsilon/dx$ is still positive, so that beyond the point where $dT/dx = 0$, ϵ continues to increase but T decreases and never attains T_m. This is not a physically satisfactory form for the reaction zone, for it does not have the essential feature of the actual zone, that the temperature rises to the value corresponding to complete reaction. For large M the solution has the property that $d\epsilon/dT$ is so small that when the solution has reached $T = T_m$, ϵ has not attained the limit value 1. This solution, too, is not physically satisfactory. There is one value of M for which the corresponding solution $\epsilon = 1$ at $T = T_m$. This solution is the true reaction zone which would be set up in steady burning (if our equations were exact descriptions of the phenomena).

Steps of 100° C have been found convenient for the integration of the differential equation. If the solution is started from the hot side of the flame, the first approximation is used until a stage has been reached at which it is possible to use the differential equation. There are points in favour of each end as the commencement of integration, but for case (II), for which our numerical computations were carried out, it has been found better to start at the low temperature side.

The advantage of numerical integration, finding M by trial and error, is that it can be used whatever the form of $d\epsilon/dT$. The only practical point to be mentioned is that it is essential to work the calculation with T as the independent variable; if ϵ were used as the independent variable slight errors in guessing T for the n arc would lead to large errors in $dT/d\epsilon$ (because of the presence of $e^{-A/RT}$), and the arc would have to be reworked a large number of times.

6. CHOICE OF THERMAL CONDUCTIVITY

In the theory of § 2, λ has been taken as a constant. In numerical integrations on the burning of nitrate esters it has been found that a simple mean value gives practically the same rate of burning as the actual conductivity which varies through the reaction zone.

The thermal conductivity is a function of ϵ and T. For low pressures it is independent of the pressure. It is usual to represent the temperature dependence by

$$\lambda(T)/\lambda(273^\circ\,\mathrm{K}) = \frac{(273+C')}{(T+C')}\frac{(T)^{\frac{3}{2}}}{(273)}, \tag{54}$$

where C' depends only on the nature of the gas, and for nitrogen and gases of similar molecular weight is about 100. The agreement with the observations is better at high temperatures than at low. Tests with a λ depending on temperature in this manner ($C' = 100$) showed that this was equivalent to a constant conductivity whose value was that appropriate to a temperature somewhat above the middle of the temperature range encountered in the flame zone.

The conductivity at any point depends on the composition of the gas at that point, that is, on ϵ. The reactant has a higher molecular weight than the final products, and its conductivity is less. For organic vapours such as benzene, λ at 0° C is about $2{\cdot}5 \times 10^{-5}$ cal./sq.cm./sec. per (° C/cm.), which is about half that of the final products at the same temperature. Assuming $C' = 100$ this would make their λ (at 2000° K) $= 0{\cdot}92 \times 10^{-4}$. Numerical solutions have shown that it is sufficiently accurate to take an average of the initial and final conductivities, when calculating rates of burning by using a theory with constant λ. Of course the (ϵ, T, x) relations are not reproduced by a constant conductivity, but this is usually not of any great importance, and the error is not large.

7. DISCUSSION

The chief result of the foregoing analysis is that we have obtained formulae which give the flame velocities in terms of the reaction velocity constants, the specific heat and heat conductivity for the three simple cases considered. These cases were: a unimolecular reaction with velocity given by equation (29), a bimolecular reaction given by equation (40), and a quasi-bimolecular reaction given by equation (23), the effect of diffusion being neglected compared with that of heat conductivity in all cases. The actual formulae are only approximations, but the approximation is one in which higher order terms are neglected in a purely mathematical method of solving the accurate physical equations. The error due to this approximation in some actual cases solved exactly by numerical integration has been found to be from 5 to 15 %. This is negligible compared with the degree of uncertainty in any initial data at the present stage.

The analysis also provides the dependence of the composition and temperature on the co-ordinate perpendicular to the flame front, to a corresponding degree of accuracy. This would give any conventional measure of the flame thickness which it is desired to make. 'Flame thickness' always depends on an arbitrary convention, since theoretically the variations of composition and temperature extend an infinite distance. The analysis shows that all particulars of the flame structure can be found to about the above degree of accuracy. If the same models are considered but with more detailed knowledge of certain properties the problem could still be solved by direct numerical integration coupled with a trial and error process for the velocity. Hence for such simple reaction systems there are no uncertainties in the theory, only detailed difficulties in the numerical work.

The qualitative aspect of the results can be summed up by stating that the mass flame velocity is proportional to the product of the square roots of the heat conductivity, the density, the reciprocal of the specific heat, and the reaction velocity near the final temperature. Hence if the homogeneous reaction velocity varies as (pressure)^n the mass flame velocity will vary as (pressure)$^{\frac{1}{2}(n+1)}$, giving $p^{\frac{1}{2}}$ for a unimolecular reaction or p for a bimolecular reaction.

In conclusion we may restate our general view that there are no real subtleties in the quantitative theory of flames but there are very great mathematical complexities.

This work was carried out for the Ministry of Supply during 1941; the authors wish to thank Professor Sir John Lennard-Jones, F.R.S., for his helpful advice and encouragement, both at that time and later. They are also indebted to the Chief Scientist, Ministry of Supply, for permission to publish.

APPENDIX

Validity of the 'local temperature' in the reaction zone

We assume that the molecules are activated by collisions with other molecules, and that the rate of activation at any point depends on the temperature *at that point only,* just as if the gas were in a large container at that uniform temperature. In other words, we assume that the mean free path of the molecules is much smaller than the thickness of the effective reaction zone. The Maxwell mean free path is

$$\bar{l} = \frac{1}{\pi 2^{\frac{1}{2}}(\text{diameter of molecule})^2\,(\text{no. of molecules per cm.}^3)}.$$

The diameter of the molecule is of order 2 to 3 Å. Let P atmospheres be the pressure. Taking 2000° K as the temperature in the most important zone of a typical flame, we find that $\bar{l}$ is of the order $10^{-4}/P$ cm. Hence, for our assumption to be valid, the effective breadth of the reaction zone must be greater than $10^{-3}/P$ cm.

REFERENCES

Jahnke, E. & Emde, F. 1933 *Funktiontafeln,* 2nd edition.
Lewis, B. & von Elbe, G. 1934 *J. Chem. Phys.* 2, 283.
Zeldowitsch, J. B. & Frank-Kamenetzki, D. A. 1938 *Acta Physicochim. U.S.S.R.* 9, 341.

Starting potentials of electrodeless discharges

By E. W. B. Gill and A. von Engel
Electrical and Clarendon Laboratories, Oxford University

(*Communicated by Lord Cherwell, F.R.S.—Received* 1 *October* 1948)

In a previous paper the starting process of electrodeless discharges in high frequency electric fields at very low pressure was shown to be controlled by the emission of secondary electrons released by oscillating electrons which hit the walls with sufficient speed. This explained why the starting field strength was found to be independent of the nature of the gas and to rise suddenly to infinity when the wave-length exceeds a critical value—the cut-off—which depends on the size of the vessel and the substance of its inner surface.

This paper records measurements of the starting field which have been carried out at pressures between about 0·2 and 350 mm. Hg in hydrogen, nitrogen, neon and some in deuterium and helium covering a range of wave-lengths between 5 and 2000 m. in vessels of different sizes and types of glass.

The results show that with increase in pressure the nature of the gas becomes of importance. In the above-mentioned gases the starting field rises slowly and continuously with the wave-length up to a region where a kind of cut-off occurs which is due to the growing amplitude of electron oscillation and the corresponding rise of wall losses. In nitrogen and hydrogen and deuterium sharp cut-offs persist up to the highest pressure used. In the rare gases these discontinuities occur only below about 0·2 mm. Hg in Ne and 0·5 mm. Hg in He, while at larger pressure the cut-off extends over a considerable range of wave-lengths. The different behaviour of molecular and rare gases seems to be caused by narrow and wide electron energy distributions respectively.

The cut-off occurs when the amplitude of oscillation of the electrons becomes equal to the length of the tube. From values of the cut-off and the associated fields the electron drift velocity has been derived. The results are in general accord with the known data which are only available for small fields. Above the cut-off a strong field is necessary to start a discharge in hydrogen and nitrogen. In hydrogen the starting field decreases with increasing wave-length and it is shown that this decrease is due to positive molecular ions producing electrons probably by impinging on the walls. In nitrogen the picture is more complex. In deuterium the positive molecular ion is effective; its mobility is found to be about one-half of that of hydrogen.

In neon at higher pressures and at wave-lengths below 100 m. the starting field rises with decreasing wave-length; this is because the electrons collide too seldom with gas atoms to attain final random velocity.

The only gas investigated at pressures comparable to one atmosphere was neon. Even at these pressures a kind of cut-off occurs. At one atmosphere the starting field is about 500 V/cm., by extrapolation it would be 15,000 V/cm. in hydrogen and of the order of 2000 V/cm. in helium for wave-lengths below the cut-off.

1. INTRODUCTION

Although many papers have been published on electrodeless discharges, the majority deal with the properties of the discharge after it has started. The present paper, which is an extension of the authors' work at extremely low pressures (Gill & von Engel 1948), is concerned solely with the potentials necessary to start the discharge. This simplifies the problem considerably because in the initial growth of current before a discharge occurs space charge effects are negligible; whereas, after the discharge starts, they modify the electric field and add much to the complexity. Other recent papers on high frequency starting potentials are by Thomson (1937); Townsend & Gill (1938); Nieserwanger, Holmes & Weissler (1946); Chenot (1947) and Hale (1948).

The notation to be used is:

λ = wave-length in metres of applied high frequency field,
$\omega = 2\pi \times$ frequency of the above,
X = peak value in V/cm. of starting-field strength,
p = pressure of the gas in millimetres of mercury or in μ ($1\mu = 1 \times 10^{-3}$ mm. Hg).

2. EXPERIMENTAL DETAILS

The discharge tubes were of cylindrical cross-section with flat ends and the main experiments were done with three tubes. Two of these were of hard glass of internal lengths between the flat ends of 3·3 and 1·7 cm. respectively, while the third was of soft glass of internal length 3·55 cm. The diameters of all three were about 7 cm.

The high frequency potential differences were applied between two large circular plates of diameter 20 cm., placed parallel and close to the flat ends so that the field was always uniform. They were the same as were used in our above-mentioned experiments, at extremely low pressures and the circuits and method of measuring the sparking potentials were exactly the same as described there. The gases, hydrogen, nitrogen, neon and helium were supplied by the British Oxygen Company and termed spectroscopically pure but the neon was stated to have a trace of helium. Our own observations from time to time with a spectroscope showed that this claim for purity was justified. Pure deuterium was produced in this laboratory by electrolysis of heavy water. The lower pressures were read by a McLeod gauge, the higher by an oil gauge and a liquid air trap was always interposed between the gauges and pumps and the discharge tube.

3. RESULTS OF MEASUREMENTS

The sparking potential in a given tube depends on both the pressure and the wave-length used and the most convenient method of showing the results is to plot the peak value of the field strength X required to start the discharge against the wave-length λ for a constant pressure p.

A series of curves for different pressures in hydrogen, nitrogen and neon are given in figures 1, 2, 3, all for the soft glass tube.

The hydrogen was studied in most detail. When the general shape of the curves was found, only sufficient curves were obtained for the other gases to see how they conformed to this general shape.

It will be seen that the main features of the curves for the three gases are the same though there are certain fundamental differences and of course the magnitudes involved differ for each gas. In all cases at the shortest wave used, namely about 5 m., X is low. As the wave-length is increased the pressure being kept constant, X rises slightly but at a certain wave-length there is a discontinuity and X rises very considerably. The wave-length at which this occurs will be called the 'cut-off' and denoted by λ_c. It is at and beyond the cut-off that differences appear in the behaviour of the three gases. The curves for hydrogen show that for each pressure X rises discontinuously up to a maximum at the cut-off. A further increase in λ for pressures up to about 1 mm. Hg produces a steady decrease in X.

For nitrogen, except at low pressure, X rises first discontinuously and then steadily so that the maximum occurs at a wave-length which is well beyond the cut-off;

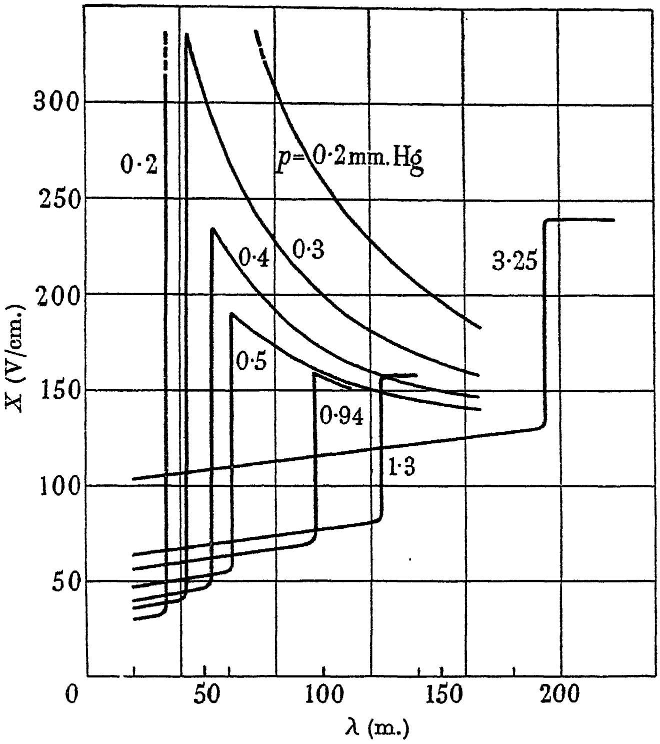


Figure 1. Field strength as a function of wave-length for a flat ended cylindrical tube of 3·55 cm. internal length, filled with pure hydrogen at different pressures.

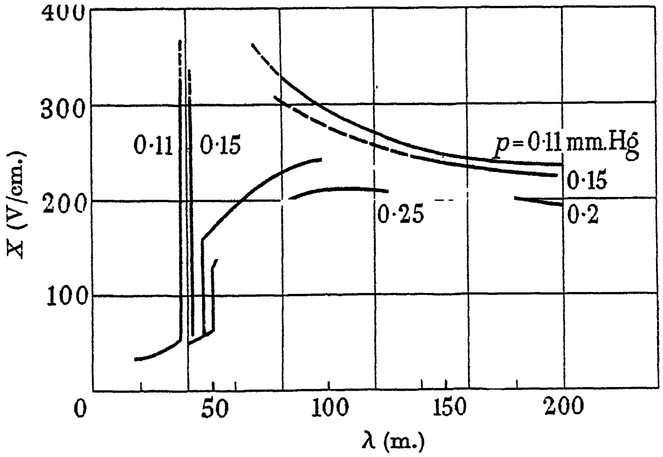


Figure 2. Field strength as a function of wave-length for a flat ended cylindrical tube of 3·55 cm. internal length, filled with pure nitrogen at different pressures.

while for neon, again except at the lowest pressures, there is no abrupt rise in X anywhere though there is a definite rise from the low values of X at small λ which extends over a large range of λ; the spread apparently increases with increasing pressure.

The curves for hydrogen show that as the pressure is increased from 0·3 to 0·94 mm. Hg the rise in X at cut-off continually decreases and it was thought at first that a further increase in pressure would reduce this rise to zero, that is the cut-off would become non-existent. Experiments at still higher pressures proved this was not the case, but that with increasing pressure this rise reached a minimum and then increased again; it seems likely that in hydrogen, whatever the pressure, there will always be a cut-off. The curve for $p = 3{\cdot}25$ mm. Hg shows this further rise; it also appears that beyond the cut-off X becomes constant at large pressures.

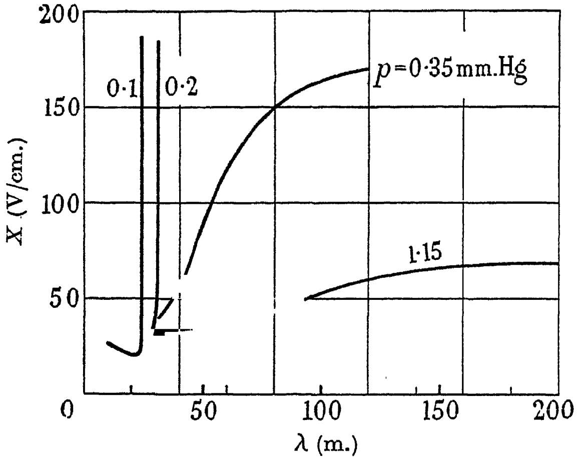


FIGURE 3. Field strength as a function of wave-length for a flat ended cylindrical tube of 3·55 cm. internal length, filled with pure neon at different pressures.

The earlier observations (Gill & von Engel 1948) at very low pressure showed similar cut-offs and the expression 'cut-off' was there strictly correct because for wave-lengths longer than the cut-off no discharge could be obtained however large the applied field. With the pressures now considered a discharge can be obtained beyond the cut-off though much larger values of X are required than just below it. For convenience the expression 'cut-off' is still applied to the wave-length where the discontinuity occurs.

It might be thought on analogy with the curve for 0·2 mm. Hg—according to which an enormous X is required for $\lambda = 40$ but only $X \approx 190$ for $\lambda = 160$—that the value of X for say 3μ would also become reasonably small for some long wave-length. This, however, is incorrect. Inspection of the values of X at $\lambda = 160$ for the pressures 0·5, 0·4, 0·3, 0·2 mm. shows that the curves are opening out more and more rapidly as p decreases and thus X will approximate to infinity for pressures of a few μ. The effect of altering the length of the tube but keeping the diameter the same is shown by the set of curves in figure 4 for hydrogen in the tube 1·8 cm. long. Comparison with the curves in figure 1, for the tube 3·55 cm. long shows that for the same pressure the general shape is the same but X is considerably larger and the cut-off wave-length in the shorter vessel smaller. Although X is larger for the short tube than for the longer, the total voltage across it of $1\cdot8X$ is somewhat smaller than the total voltage $3\cdot55X$ across the larger.

Some experiments were also done with two tubes of the same length but different diameters. Again the general shape of the curves was the same but an increase in diameter decreased X and increased λ_c. The change in diameter has a relative smaller effect on X and λ_c than a corresponding change in tube length.

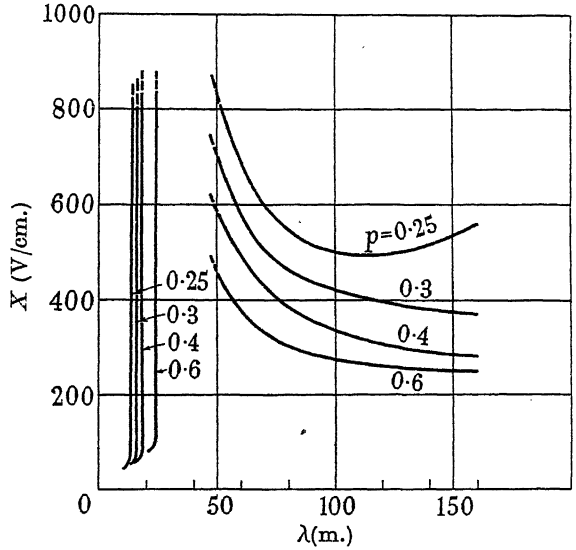


FIGURE 4. Field strength as a function of wave-length for a flat ended cylindrical tube of 1·8 cm. internal length filled with pure hydrogen at different pressures.

It is worth remarking that below the cut-off, whatever the gas, considerable accuracy of measurement is possible; successive readings of X rarely vary by even as much as 1 %. Above the cut-off this is different. In a series of readings of X for the same λ the first reading is always the highest and readings after that are grouped fairly closely round a lower value. As will be seen later, it is probable that above the cut-off bombardment of the sides of the tube has an effect, but has little effect below the cut-off; and the condition of the sides is likely to be affected by the first discharge. This inaccuracy above cut-off is most apparent with neon and least with hydrogen.

Quite independently of this, experiments get more difficult at long wave-lengths because the discharge is so feeble that it is difficult to see at what voltage it starts. At the shorter wave-lengths not only is there plenty of light but the discharge current is large enough to cause considerable detuning in its circuit accompanied by an instantaneous drop in the electrostatic voltmeter.

4. SIMPLE THEORY UP TO THE CUT-OFF

The initiation of an electrodeless discharge requires that the few electrons normally present in the gas shall be multiplied many million times by the action of the alternating field. By far the most effective kind of multiplication is usually ionization of the gas by the electrons; other causes are secondary emission from the sides of the vessel and to a much less extent effects due to the positive ions or radiation. But while this increase is going on, there is simultaneously a decrease because some electrons are reaching the sides and are lost; there is also a loss by recombination which becomes important at higher pressures. The essential condition for a discharge to start is that the rate of production of new electrons must always exceed the rate of loss. This rule must operate right from the start, when the field is first applied, and the initial number of electrons is so small that, even after it has grown by a factor of thousands, the field due to space charges is still negligible. The suggestion that very low starting potentials are due in some way to space charge effects, based on the fact that when the discharge once starts a much smaller potential will maintain it, is untenable. At the start the alternating field alone must produce the increase whatever help it may perhaps get at a later stage from space charge fields.

An exception to the rule that gas ionization is the preponderating factor is the case of extremely low pressures discussed in the above-mentioned paper by the authors. At pressures of the order 1μ the majority of electrons traverse the tube without any collision with gas molecules and their losses to the ends and sides can only be compensated by secondary emission from the glass which is here the most important factor.

It is a very different thing when the tube contains gas at an appreciable pressure, the electrons have many more collisions with gas molecules than with the walls. There is a stronger reason still why secondary emission plays a very small part. For secondary emission to be effective it is necessary, not only for the primary electrons to hit the ends with a considerable velocity, but the reversing field must move the secondaries in the opposite direction. This is possible at low pressures because electrons moving in the alternating field do not come to rest when the field passes through zero, but owing to their initial velocity move on for a time against the field which is in the proper direction to remove the secondaries (Gill & von Engel 1948). But if the pressure is sufficiently large the velocity of the electrons is practically in phase with the force, and it is impossible to get simultaneously the necessary velocity of impact and the right field direction for removing the secondaries.

If secondary emission and other minor causes of ionization are ruled out and only ionization by collision is operative cut-offs must occur, since electrons reaching the ends are lost. If the pressure is not too low the drift velocity of the electrons depends on the electric force. No simple relationship exists between the velocity and the force but, except for small forces, the velocity in most gases is approximately proportional to the force. For simplicity therefore the drift velocity will be taken as $kX \sin \omega t$, where k is a constant, i.e.

$$\frac{dx}{dt} = kX \sin \omega t. \tag{1}$$

Integration shows that the total amplitude of oscillation of the electron is $2kX/\omega$ while its maximum velocity is kX. Now for ionization to occur the drift velocity must reach a certain velocity, say v, so that kX cannot be less than this value. Therefore the total amplitude of oscillation is $2v/\omega$. The greatest total amplitude is the length of the tube, so if d is this length, the frequency of the limiting wavelength λ_c is given by

$$d = 2v/\omega, \tag{2}$$

where it will be noted, v is the velocity at the peak, i.e. when the force is X. Although the above is a considerable oversimplication, the equation (2) when everything is taken into account still remains substantially correct.

The simplest picture is therefore that the electrons oscillate to and fro having a maximum velocity proportional to X at the centre of oscillation. If this drift velocity is large enough the random velocity is large enough to produce ionization at the centre and the new electrons join in the oscillation and in turn ionize so that a constant increase occurs. After an ionizing (inelastic) collision the electrons have not quite the same centre of oscillation as the original ones because speed is lost at a collision, so that after it the electrons initially have not the speed appropriate to the centre of the original oscillation. They pick up speed quickly and oscillate about a slightly displaced centre, so that as time goes on the increasing stream of electrons occupy more and more length of tube. The tube is filled with such streams centred on the places electrons happened to occupy when the force was turned on. Any of the electrons reaching the ends (or sides) are lost and if the rate of ionization exceeds the rate of loss a discharge will occur. The electrons normally present in any volume of gas are distributed at random points, and when the alternating force is applied each of these points becomes the centre of a constantly increasing and slowly expanding stream provided that the force is sufficient to give ionizing speed at the centre. If the amplitude required for this central velocity is small compared to the length of the tube the loss of electrons at the ends will be small and very little ionization at the centre will suffice to exceed this loss and cause a discharge. For larger amplitudes the losses go up and more ionization is required, so that X has to go up; and the maximum permissible amplitude is the tube length.

These considerations account for the slow rise in X as λ is increased up to the cut-off because the amplitude to get a specified maximum velocity increases with and is proportional to λ.

This simple picture requires considerable expansion. The electrons, because of their collisions with gas molecules, get a random velocity which increases with X, and instead of electrons oscillating harmonically, we have to consider a swarm of particles moving in all directions, but the swarm as a whole moving backwards and forwards harmonically. The average velocity of the swarm is known as the 'drift velocity' and the random velocity is considerably greater than the drift velocity. The random velocity at any moment is not of course the same for each electron, but there is a velocity distribution.

Consider such a swarm which at the time when $X \sin \omega t$ is zero occupies a small volume. It moves down the tube, the drift velocity and the random velocity both varying as it proceeds. Both reach a maximum close to the time when $X \sin \omega t$ passes its maximum and thereafter decrease till the drift velocity becomes zero and the swarm then turns back. The question is: does the swarm keep together in this process or does the random velocity in all directions cause an appreciable spreading out in the direction of the field.

If the electrons keep together, a small increase in wave-length above that which just corresponds to the tube length will take them all on to the glass and a clean cut-off will occur. But if they are much spread out many will be left behind in the gas when reversal occurs and an increased X may give sufficient ionization to overcome the loss of the remainder. There will certainly be a noticeable increase in X when the majority reach the ends but no clean cut-off. One can be slightly more precise here. A rise in X increases the amplitude of oscillation and this in turn increases the losses. Obviously the condition for a discharge to occur is that the rate of increase of ionization with X must exceed the increase in the rate of loss at a given λ. It should also be remarked that, although the random velocity depends on the drift velocity, the connexion is not a linear one and in particular when $X \sin\omega t$ passes through zero and the drift velocity is zero the random velocity is not zero, which means that the losses to the ends are somewhat more than on the simple idea of electrons oscillating harmonically.

To follow this out in detail is beyond the scope of this paper which is concerned with experimental facts and their general explanation; but the results appear to indicate that the stream of electrons keeps well together in hydrogen, not very well in nitrogen, and extremely badly in neon. The work of Townsend contains certain considerations which are suggestive. The spreading out is likely to depend on the ratio of the random velocity v_r to the drift velocity v. If the ratio v_r/v is large one would expect a large spread and vice versa. Townsend found that this ratio depends on X/p; and that the ratio becomes smaller the larger X/p. Thus one would expect a small spread and a sharp cut-off for the lower pressures, and this is true.

5. Calculation of electron mobilities

The fact that at a wave-length just below the cut-off the amplitude of oscillation equals the length of the tube enables the mobilities of the electrons to be measured. From our point of view the calculation is a test of general theory because these mobilities are known, having been measured by various methods by previous investigators. Equation (2) indicated that if the velocity can be taken as proportional to X then the frequency $\omega/2\pi$ of the cut-off wave-length λ_c in a tube of length d is given by $\omega = 2v/d$, where v is the maximum drift velocity. λ_c therefore gives v which occurs at the maximum of $X \sin\omega t$, i.e. corresponds to the X of this value of λ_c. This assumption that v is approximately proportional to X is in accordance with published results (Bradbury & Nielsen 1936; Nielsen 1936). For a given X this occurs when p is sufficiently large. If p is reduced making X/p large the mean free path gets longer and the mean velocity between collisions is proportional to $\sqrt{X}$. Thus for the high values of X/p

$$\frac{dx}{dt} = k_1\sqrt{(X \sin\omega t)}. \tag{3}$$

This integration involves γ-functions and equation (2) must be replaced by

$$\frac{d}{1{\cdot}2} = \frac{2v}{\omega}. \tag{4}$$

where $v = k_1\sqrt{X}$.

This treatment for high values of X is not of course rigid because $X \sin \omega t$ passes through all values from 0 to $\pm X$, so that there is a region of the integration where the square root law is incorrect. This region, however, is for the small values of $X \sin \omega t$, where the distance covered is small and the major part of the path is determined by the square root law.

It is important to realize that the preceding argument can only apply if the pressure is high enough for the velocity to be in phase with the force. At the other extreme is the case investigated in our original research at extremely low pressures. Here with electrons moving practically in a vacuum the velocity cannot be attributed to a force X but instead to a constant acceleration. In any case here secondary electrons are of prime importance and at the cut-off wave-length the velocity is out of phase with the force. As the pressure is increased from this low value the secondary electrons become progressively less important, the phase angle decreases and the electrons tend towards a limiting velocity approximately proportional to X. What pressure range this intermediate condition covers, and below what value of X/p the velocity v can be calculated accurately is difficult to say and must be got at indirectly. For instance if the results calculated from tubes of different lengths agree, the answer is probably accurate.

Some experiments were made to see at what pressure the influence and secondary emission at the starting field became negligible. Two tubes of identical dimensions were used and the flat ends of one tube were coated with palladium black which has a very poor secondary emission compared to glass. Using hydrogen it was found that at 0·2 mm. Hg the value of X for the coated tube was some 30 % higher than for the uncoated tube; with increasing pressure the difference decreased and at about 0·5 mm. Hg, it disappeared for all value of λ, up to the cut-off. Hence secondary emission from the glass is completely ineffective in H_2 at pressures exceeding 0·5 mm. Hg.

This pressure range of uncertainty being dependent on the mean free paths depends on what gas is used; and it is of interest that with the larger vessel, the cut-offs for hydrogen and neon are practically identical up to about 0·4 mm. Hg, but for nitrogen they very soon deviate as p is increased above 10μ. This shows that for the fields used the mean free paths are much longer in hydrogen and neon than in nitrogen and that in these gases up to 0·4 mm. Hg things are the same as at 1μ where secondary emission is the controlling factor. Evidently in a range where the fundamental processes are independent of the nature of the gas no information can be obtained on velocities appropriate to the gas.

6. DISCUSSION OF THE RESULTS

Hydrogen

Two tubes were used of internal lengths 3·55 and 1·7 cm. and equation (2) gives for these maximum velocities $23{\cdot}6 \times 10^8/\lambda_c$ and $16{\cdot}2 \times 10^8/\lambda_c$ cm./sec. respectively, λ_c being the cut-off wave-length in metres. These velocities correspond to the values of X_c just below cut-off. It is over this value of X_c that some uncertainty arises. The curve X-λ rises steadily with increasing λ over a long range but just before the cut-off it bends upwards before the discontinuity. This bend, due in all probability, even in hydrogen, to a small velocity distribution was disregarded and X taken at the value it would have for the straight curve at the cut-off wave-length. In figure 5 λ_c and the corresponding values of X_c are plotted against p which refers to the 3·55 cm. vessel and from this curve the results in table 1 were obtained.

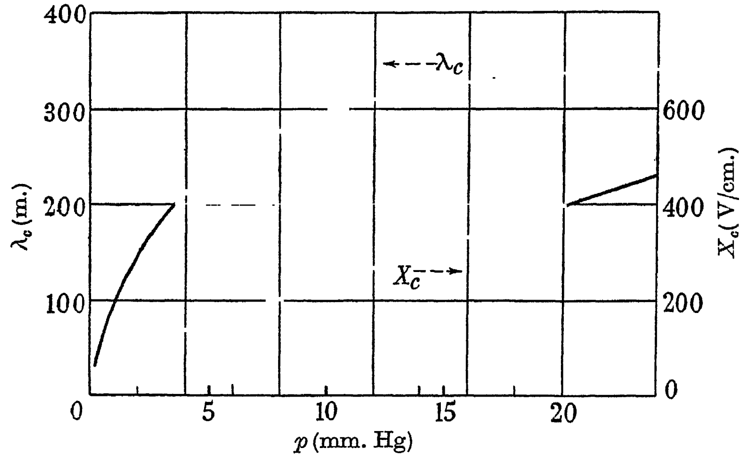


FIGURE 5. Cut-off wave-length and field strength as a function of gas pressure for a hydrogen filled vessel of 3·55 cm. internal length and 7 cm. diameter.

TABLE 1

p (mm.)	X(V/cm.)	X/p	v (cm./sec.)	$v/(X/p)$
24	470	19·6	$9{\cdot}8 \times 10^6$	5×10^5
16	340	21·2	10·2	4·8
10	242	24·2	11·3	4·67
6	136	34	16·0	4·7
3	117	39	17·9	4·6
2	100	50	24·4	4·86
1	65	65	33·6	5·15

Below 1 mm. it is doubtful if the formula can be safely used; if it were correct $p = 0{\cdot}6$ gives $v = 35 \times 10^6$ for $X/p = 92$, and $p = 0{\cdot}2$ gives $v = 99 \times 10^6$ for $X/p = 185$. Over the safe range v, as the last column shows, is practically proportional to X/p, so the linear law was justified. Similar experiments with the smaller tube gave results differing by only a few per cent from the above, although at the same pressure the values of λ_c and X_c at cut-off were widely different. Roughly speaking for identical pressure λ_c for the longer tube was about $2\frac{1}{2}$ times that for the shorter and the X_c was slightly smaller.

Previous experimenters have not been able to determine v for values of X/p above 20, and merely for constant field. It is only possible therefore to compare our results with theirs for the lower values of X/p. Bradbury & Nielsen (1936), for $X/p = 20$, finds $v = 7 \times 10^6$, so that our result is certainly of the right order. No claim is made to high accuracy. The most accurate figure is the cut-off wave-length; it can be got easily to 1 %. X_c is more open to doubt, it is not more than 1 or 2 % out for the highest pressures but may be some 5 or 6 % inaccurate at the lower. The length of the tubes introduces a further error of several per cent as the ends were not truly plane and parallel. Nevertheless the fact that the results for the two tubes agreed within a few per cent suggests the error is only of this order.

Nitrogen

The uncertainty over the cut-off X was more pronounced than in hydrogen as there was considerably more bending round of the curve before the final discontinuity. The dependence of λ_c and the most likely values of X_c on the pressure is shown in figure 6. From this come the values shown in table 2; v_1 and v_2 are the drift velocities on the assumption that they are proportional to X and $\sqrt{X}$ respectively.

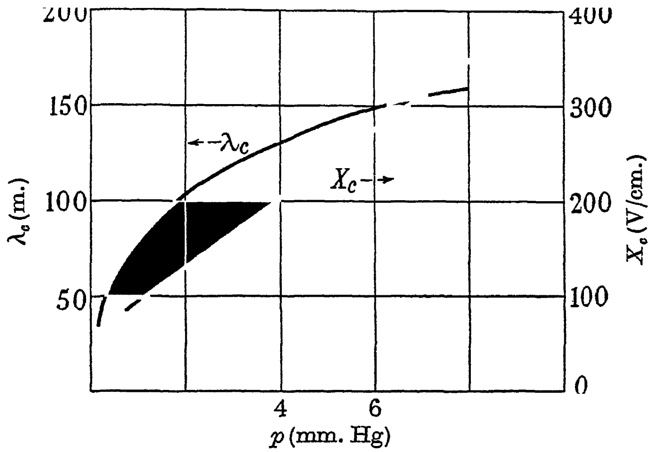


FIGURE 6. Cut-off wave-length and corresponding field strength as a function of pressure for a nitrogen filled vessel of length 3·55 and 7 cm. diameter.

TABLE 2

X/p	v_1(cm./sec.)	$v_1/(X/p)$	v_2(cm./sec.)	$v_2/\sqrt{(X/p)}$
50	$23{\cdot}2\times10^6$	$0{\cdot}46\times10^6$	$19{\cdot}4\times10^6$	$2{\cdot}75\times10^6$
60	28·9	0·48	24	3·08
70	33	0·47	27·5	3·29
80	36·9	0·46	30·8	3·45
90	40·2	0·45	33·6	3·53
100	43·5	0·44	36·3	3·63
150	53·2	0·39	44·5	3·6
200	61	0·31	51	3·6
300	70·7	0·24	59	3·36
400	80	0·20	66·7	3·33
500	88·2	0·18	74	3·33

The third column shows that v_1 is proportional to X/p for values of from 50 to about 100 while the last shows v_2 is proportional to $\sqrt{(X/p)}$ for X/p between 300 and 500.

v_1 therefore gives the true value of the velocity for $X/p = 50$ to 100, v_2 gives it for $X/p = 300$ to 500 and for X/p in range 100 to 300 the true value is between v_1 and v_2. The result for $X/p = 150$ came from a pressure of 0·5 mm. and the higher values of X/p involved pressures lower than this so that they are in the doubtful region. Experiments with the shorter tube, as in the case of hydrogen, agreed with the longer tube within errors of experiment. The biggest value of X/p used by Nielsen was 20 and the drift velocity was $7{\cdot}5 \times 10^6$. From the above table our result would be $9{\cdot}2 \times 10^6$ so that the order is correct.

Neon

Above 0·25 mm. Hg no discontinuity is obtained and it is not considered here that any velocity calculation is possible. Further, it is only at the very lowest pressures that even an inaccurate determination of X is possible, and here secondary emission and phase angles are likely to vitiate the results. Some experiments at high pressures give an idea, however, of the order of the velocity for very low values of X/p.

Using the smaller discharge tube and a pressure of 350 mm. the X-λ curve (the dotted curve in figure 8) shows first a small steady rise followed by a considerable rise. Above $\lambda = 1100$, X remained constant. This may be interpreted as follows: there is evidently a considerable velocity distribution among the stream arriving at the ends of the tube. The first abnormal increase in X near $\lambda = 700$ is due to the fastest electrons in the stream, somewhere about $\lambda = 800$ to 900 the main body have arrived and in the region of $\lambda = 1000$ the slowest. On this basis one gets the approximate results. Fastest electrons: when $X/p = 0{\cdot}63$, $v = 2{\cdot}4 \times 10^6$. Main body in range $X/p = 0{\cdot}69$, $v = 2{\cdot}2 \times 10^6$ to $X/p = 0{\cdot}75$, $v = 1{\cdot}9 \times 10^6$. Slowest $X/p = 0{\cdot}83$, $v = 1{\cdot}7 \times 10^6$. Similar calculations at $p = 250$ mm. gave lower values of v so that the results are not very reliable. The corresponding value as measured by Nielsen (1936) for $X/p = 0{\cdot}7$ is $1{\cdot}1 \times 10^6$ cm./sec.

Helium

Preliminary experiments show that the character of the curves is similar to the ones obtained with neon; discontinuous cut-offs have been observed up to larger values of p (order 0·5 mm.) than with neon. Extrapolating from measurements at higher pressure the starting field of helium at 1 atmosphere appears to be about four times higher than that for neon (see § 9).

7. GENERAL DISCUSSION OF CURVES UP TO THE CUT-OFF

What mainly emerges—at any rate in tubes of the sizes used—is the controlling effect of losses to the glass, chiefly to the ends but partly also to the sides. If the tube length is increased, X decreases at all wave-lengths showing that the losses become smaller; the same is true if the diameter is increased. This decrease in X results in λ_c being increased. No simple relation has been discovered between X, p, λ and the length of the tube except at the cut-off. Here, provided that v is proportional to X/p, $v = \text{constant} \times X_c/p$. From (2) at the cut-off v is proportional to d/λ_c so that

$$\frac{X_c \lambda_c}{pd} = \text{constant}. \tag{5}$$

Hence in any tube the cut-offs obey the law $X_c\lambda_c/p = \text{constant}$ and in different tubes this contant is proportional to the length. The experimental results all indicate that the cut-off X, for any tube, plotted against p gives a straight line cutting the X-axis at a positive value. If this straight line is correct to high values of p a simple connexion between cut-off wave-length and p is obtained, viz.

$$\lambda_c\left(a+\frac{b}{p}\right) = \text{constant}. \tag{6}$$

a and b are constants showing that λ_c tends to a fixed value at high pressures, as is indicated by the curves (figures 5 and 6).

8. Discharges at wave-lengths exceeding the cut-off

At the cut-off the starting field rises considerably, but except at the lowest pressures a discharge can still be started at wave-lengths exceeding the cut-off though a strong field is necessary.

On the preceding considerations all the electrons will be lost to the walls at these wave-lengths and field strengths. For instance in hydrogen at 0·3 mm. Hg at the cut-off $\lambda = 43$ with $X = 36$, the amplitude of the electron oscillation just fills the tube; however, at $\lambda = 44$ a discharge occurs if $X = 340$, i.e. where the amplitude is nearly ten times the tube length, so that every electron crosses the tube and is lost at the ends in a fraction of a half period.

Some other mode of ionization must have come into operation and the clue as to what it is is the shape of the X-λ curves for constant p above cut-off. The hydrogen curves, figure 1, show the phenomenon most clearly. What is most striking is that a slight increase in wave-length just beyond the cut-off at the lower pressures produces a large or very large reduction in X. That is with increasing wave-length the discharge starts more and more easily or in other words the ionizing agency, whatever it is, becomes more and more effective with increasing wave-length. This rules out radiations, metastable atoms and the like and suggests as the most simple agent a material charged particle which by reason of its inertia will act differently for different wave-lengths.

Without at first speculating on its nature it is sufficient to assume merely that it has a charge e (either positive or negative) and a mass M. It will also be assumed that its motion is retarded by a frictional force proportional to its velocity; and the last assumption is that, whatever the wave-length, to be effective it must attain the same maximum velocity in its oscillations. This last seems reasonable for whatever the wave-length it must produce the same number of electrons and ions; and however this takes place the velocity is a determining factor. The equation of motion of the particle in an applied field $X\cos\omega t$ is

$$M\frac{d^2x}{dt^2} = eX\cos\omega t - k\frac{dx}{dt}. \tag{7}$$

Integration gives

$$\frac{dx}{dt} = -\frac{eX\sin\phi}{k}\sin(\omega t-\phi), \quad \text{where} \quad \tan\phi = -\frac{k}{M\omega} \tag{8}$$

and

$$x = \frac{eX}{\omega\sqrt{(M^2\omega^2+k^2)}}\cos(\omega t-\phi). \tag{9}$$

The maximum velocity is $$\frac{eX\sin\phi}{k} = \frac{eX}{\sqrt{(M^2\omega^2+k^2)}}. \tag{10}$$

Hence, if the maximum velocity is constant and equal to v, we have

$$e^2X^2 = v^2(M^2\omega^2+k^2). \tag{11}$$

Thus if X^2 is plotted against ω^2 the curve should be a straight line. These plots are shown in figure 7 for hydrogen at pressures of 0·2, 0·3, 0·4 and 0·5 mm. respectively and turn out to be accurately straight lines. (The complementary function in (8) was found to be negligible for the magnitudes involved.)

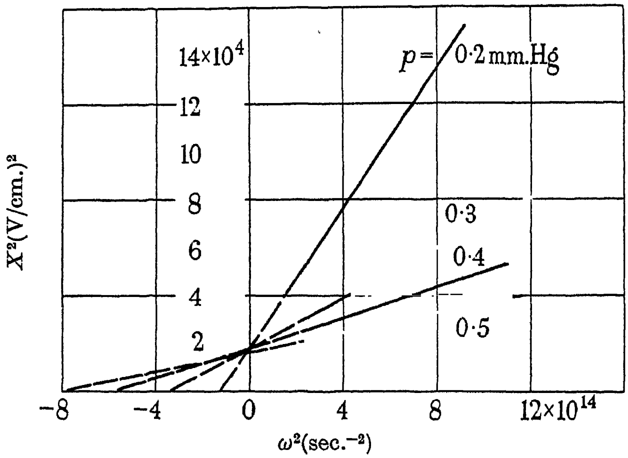


FIGURE 7. Relation between field strength and frequency beyond cut-off in hydrogen at various pressures (dashed lines: extrapolated).

The assumptions thus appear to be justified and the next step is to find out more of the nature of the oscillating particles. For any pressure p the straight lines cut the X axis at a point X_0, where

$$X_0 = \frac{vk}{e} \tag{12}$$

and cut the ω-axis at a point $-\omega_0$, where

$$\omega_0 = \frac{k}{M}. \tag{13}$$

We therefore know vk/e and k/M, but this does not give M or the velocity v without further assumptions.

The frictional constant k can be replaced by a more convenient quantity; if the charged particle moves in a constant field X, it obtains a limiting drift velocity $v = eX/k$ so that if μ is the corresponding mobility

$$\mu = \frac{e}{300k}, \tag{14}$$

so from equation (13) $$\mu = \frac{e}{300M\omega_0}. \tag{15}$$

The identification of the particles is therefore the indirect method of taking e/M for possible candidates for the position and seeing which of them has, from equation (15), a velocity agreeing with measurements of previous researchers. On this basis the particles are identified as the molecular hydrogen ions for which e/M is taken as $1{\cdot}43 \times 10^{14}$. It seems less likely to assume that the value of e/M as found is associated with a mixture of H^+ and H_3^+ because their properties would have to be independent of the pressure and also little H_3^+ is to be expected to exist at medium p.

From equation (15) the curves, figure 7, and this value of e/M we obtain the ion mobilities μ and μ_1, the corresponding velocity for 1 mm. pressure, both in cm./sec./V/cm. The fact that the calculated μ is practically proportional to the inverse pressure is a confirmation of the general theory and the mean value of μ_1 is not very different from the latest known figure of 7600 for H_2^+ (Bennet & Thomas 1942). Considering all the possible errors it would have been satisfactory if the result had merely been of the right order, so the actual agreement is striking (for earlier results see Tyndall 1938).

p	μ	μ_1
0·2	$4{\cdot}29 \times 10^4$	8600
0·3	2·51	8400
0·4	1·98	8000
0·5	1·70	8500

The maximum velocity v attained comes from the intercept X_0 on the X-axis; from (12) and (14) with X_0 in V/cm.,

$$v = \mu X_0 = \frac{X_0}{p}\mu_1. \tag{16}$$

It is interesting that the curves all seem to cut the X-axis at about the same point, the approximate value of X_0^2 being $1{\cdot}8 \times 10^4$ or $X_0 = 135$. Hence these velocities are inversely proportional to the pressure and are given by $v.p = 1{\cdot}1 \times 10^6$ cm./sec. which is in general accord with Townsend's work that ionization depends on X/p. Equation (9) shows that the amplitude of oscillation for H_2 to attain these velocities is only a fraction of the tube length. The conclusion is therefore suggested that, above the cut-off, the major part of the ionization is due to positive ions which in order to start the discharge require a maximum velocity inversely proportional to the pressure.

It is too early to say what the actual mechanism is. The possibilities are ionization of gas molecules as Townsend thought, or secondary emission from the ends or sides due to bombardment by these positives. It is very doubtful if the velocities are sufficiently high for the first of these alternatives, and a small experimental fact points to the latter. This fact is that above the cut-off results are somewhat inconsistent. The potential to produce the first discharge is always higher than for subsequent ones. It is difficult to see how the first discharge can alter the gas, but it can by bombardment affect the glass or produce a surface layer. The fact that in the range $p = 0{\cdot}2$ to $0{\cdot}5$ mm. Hg. the value X_0 at zero frequency is independent of p suggests that the processes involved must be ultimately controlled by something else, i.e. wall effects.

Electron emission from the walls due to the impact of positive ions is a very different thing from the emission due to electron impact. In the latter case the number set free depends not only on the number impinging but also on the velocity of impact which must be high. The velocity of the ions is not the only important factor for low velocity ions are quite effective. For metals one electron is released for each 10 to 100 of low velocity ions; for glass the figures are not known but we suspect they are of the same order. The effect of the positive hydrogen ions depends therefore primarily on the rate of arrival, and this in turn depends upon v.

The emission coefficient γ of glass can however be estimated from figure 1 assuming that gas ionization by electrons is effective and Townsend's equation of sparking potential holds. Again it is found that γ is of the order 10 to 100 ions per electron released.

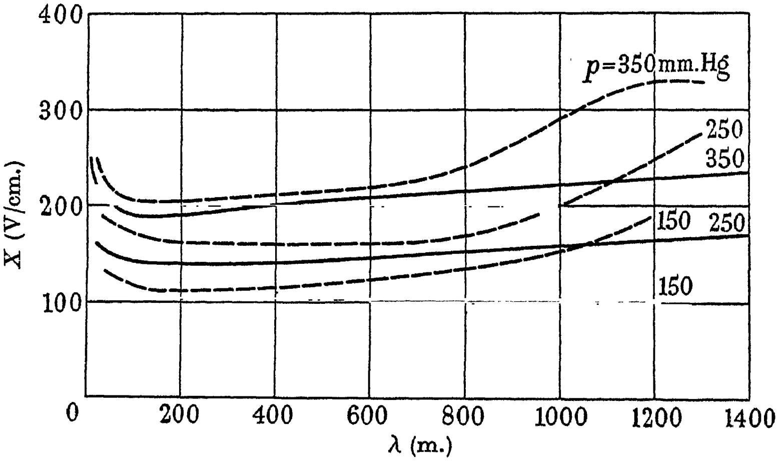


FIGURE 8. Field strength as a function of wave-length at large pressure in neon for vessels of 1·7 (dashed) and 3·55 cm. (full) internal length and 7 cm. diameter.

Because of collisions notwithstanding the small amplitude of oscillation the rate of arrival at the ends is very large, the ions acquiring a random velocity. Electrons are thus constantly set free and for half of them the field will be in the right direction to take them across the tube producing fresh electrons and ions in the gas.

Equation (8) shows that for long wave-lengths the motion of the ion is determined mainly by its mass whereas for short wave-lengths k, i.e. the mobility becomes important.

The encouraging results in hydrogen were not fulfilled for nitrogen or neon. In these gases when X^2 was plotted against ω^2 [see (11)] straight lines were not obtained with any certainty. The most likely explanation is as follows: in these gases owing to the spread of the electrons as they cross the tube a certain number are left in the gas and are not lost to the ends in each half cycle. The total ionization is therefore due not only to the positive ions but also partly to this residue of slow electrons, and the effect of the positives is therefore partly masked and the straight lines are not obtained.

9. RESULTS AT HIGHER PRESSURES

The main experiments were made with pressures of a few millimetres of mercury, but in neon the potentials required at much higher pressures were still low enough to fall in the range possible to our transmitters. Figure 8 shows results for neon for pressures 150, 250 and 350 mm. respectively. The full lines are for the tube of internal length 3·55 cm., the dashed for 1·7 cm. length. As mentioned before even at 350 mm. pressure the small tube shows a species of cut-off extending from $\lambda = 700$ to $\lambda = 1100$. The cut-off for the big tube would have been at a longer wave-length than was readily available. In any case such excessive wave-length measurements are unreliable because of the weakness of the discharge when it starts and consequent difficulty of observation.

Below the cut-off the difference between X for the small and big tube becomes proportionately less as the pressure increases. At the high pressures the losses due to recombination in the volume of the gas become increasingly important compared with the losses to the sides, and thus the size of the tube becomes of less importance except near the cut-off.

At the higher pressures for any given wave-length X becomes proportional to p; in fact a comparison of the curves for the big tube at 350 and 250 mm. shows this proportionality practically established at these pressures. For neon X/p is of the order of 0·6 at atmospheric pressure, the peak value of the starting field strength will only be about 450 V/cm. For hydrogen the corresponding figure would be of the order of 15,000 V/cm. It would not appear that at the higher pressures any ionizing effects by positive ions are necessary. The spread of the electrons is so considerable that enough are left in the gas when the force reverses to carry on the multiplication. As explained in § 4 the very low value of X/p is likely to produce a considerable spread so that, although the curves for the small vessel show a considerably increased loss to the ends at the longer wave-lengths, there are plenty of electrons lagging behind and not lost.

It will be observed that as λ is reduced below about 100 m. the value of X increases for these high pressures. It was thought at first on analogy with the ideas discussed in § 7 that this was due to the inertia of the electrons, but calculation showed that this was not so. What is most probable is that at these shorter waves the appropriate drift velocity is attained but not the corresponding random velocity. The random velocity is only attained after a considerable number of collisions, and at the shorter wave-lengths there is not enough time near the peak for numerous collisions to take place. For neon at such very low X/p the mean free path of an electron becomes very large and this reduces the number of collisions.

10. EXPERIMENTS WITH DEUTERIUM

A very few experiments have been done with deuterium. These preliminary observations appear to show that, compared with hydrogen at the same pressure, the cut-off wave-length is slightly larger and X at cut-off much the same, also that sharp cut-offs persist up to the highest pressure; above the cut-off the X-λ curves are not very different. The first-mentioned observation proves that the electron mobility in deuterium is slightly lower than in hydrogen and the second shows that the mobility of the positive deuterium ions is also lower than the mobility of molecular hydrogen ions. If the curves coincided the mobility of deuterium would be half of that of hydrogen. Actually it has been observed to be about 0·7 (Bennet & Thomas 1942).

Our thanks are due to Professor Lord Cherwell for advice and every facility for this work. Dr H. Halban and Mr R. Wilson have kindly supplied pure deuterium.

REFERENCES

Bennett, W. H. & Thomas, L. H. 1942 *Phys. Rev.* **62**, 41.
Bradbury, N. E. & Nielsen, R. A. 1936 *Phys. Rev.* **49**, 388.
Chenot, M. 1947 1re Thèse, Fac. des Sciences, Paris.
Gill, E. W. B. & von Engel, A. 1948 *Proc. Roy. Soc.* A, **192**, 446.
Hale, D. H. 1948 *Phys. Rev.* **73**, 1046.
Nielsen, R. A. 1936 *Phys. Rev.* **50**, 950.
Nieserwanger, C. R., Holmes, J. R. & Weissler, G. L. 1946 *J. Opt. Soc. Amer.* **36**, 581.
Thomson, J. 1937 *Phil. Mag.* **23**, 1.
Townsend, J. S. 1925 *Motion of electrons in gases.* Oxford University Press.
Townsend, J. S. & Gill, E. W. B. 1938 *Phil. Mag.* **26**, 290.
Tyndall, A. M. 1938 *The mobility of positive ions in gases*, p. 53, Cambridge Univ Press.

Momentum and energy diffusion in the turbulent wake of a cylinder

BY A. A. TOWNSEND, *Emmanuel College, University of Cambridge*

(*Communicated by Sir Geoffrey Taylor, F.R.S.—Received* 6 *October* 1948)

A detailed experimental investigation of the turbulent motion in the wake of a circular cylinder, 0·953 cm. diameter placed in an air-stream of velocity 1280 cm.sec.$^{-1}$, has been carried out with particular reference to those quantities determining the transport of turbulent energy and mean stream momentum. At distances of 80, 120 and 160 diameters down-stream from the cylinder, direct measurements have been made of mean flow velocity, turbulent intensity, viscous dissipation, energy diffusion, scale, and form factors of the velocity components and their spatial derivatives. These observations show that, except close to the wake centre, the flow at a point fixed with respect to the cylinder is only intermittently turbulent, due to the passage of the point of observation through jets or billows of turbulent fluid emitted from the inner wholly turbulent core of the wake. Further consideration of the results indicates that the turbulent motion within the jets is solely responsible for the turbulent transfer of momentum, while diffusion of turbulent energy and of heat is carried out by the bulk movement of the jets. Most probably, the jets are initiated by local fluctuations of pressure inside the turbulent core, and in the later stages of their development that are slowed down by adverse pressure gradients. The existence of pressure-velocity correlations of sufficient magnitude is demonstrated by using the equation for the conservation of kinetic energy in the wake, all terms of which are known excepting the one involving the pressure-velocity correlation, which is then obtained by difference. While the conception of jets of turbulent fluid is more convenient for following the physical processes in the wake, the alternative but equivalent description that the turbulent motion consists of a motion of scale small compared with the mean flow superimposed on a slower turbulent motion whose scale is large compared with the mean flow may be used. A formal explanation of this two-stage turbulent structure in terms of the Fourier representation of the velocity field is suggested, which relates the structure to the presence of a quasi-constant source of energy of nearly fixed wave-number, and to the free boundary which allows an unlimited range of wave-numbers. It is expected that this type of motion will occur in all systems of turbulent shear flow with a free boundary, such as wakes, jets and boundary layers.

INTRODUCTION

Recent developments of the statistical theory of isotropic turbulence have made possible a reasonable and consistent explanation of the phenomena attending decay of the turbulent flow produced behind a uniform mesh placed in a constant velocity air-stream. At any moment, the turbulent flow is considered to be in a statistical equilibrium, established by the continuous transfer of energy from large eddies to smaller eddies. When this process is in moving equilibrium, then predictions of the form of the energy decay law, of the spectrum function and of the correlation functions may be made in substantial agreement with experiment. This line of thought was originated by Kolmogoroff (1941), independently by von Weizsacker (1948) and Onsager (1945), and developed by Heisenberg (1948), Batchelor and the author (Batchelor & Townsend 1947, 1948; and Batchelor 1947, 1948). The full application of the theory has only been possible in the special case of uniform flow, but, if only eddies small compared with the dimensions of the mean flow system are considered, it is expected that the theory will be equally valid for steady shear flow. This has recently been confirmed, and a considerable simplification of the problem of turbulent shear flow has resulted (Townsend 1948*a*). It is still true, however, that many of the properties of shear flow are associated with the large eddies not included in the theory, and a series of measurements has been made with the purpose of measuring the turbulent transfer of energy and momentum in the wake of a circular cylinder. To interpret the results of these experiments, it has been necessary to modify very considerably the existing notion of the structure of a turbulent wake, and an attempt has been made to provide a self-consistent description of the wake capable of representing all the known properties. As yet, this description is very empirical and most qualitative, but it is believed to represent the main features of wake motion.

NOTATION

Consider the wake of a circular cylinder of diameter d, placed in an air-stream of velocity U_0. Then establish a co-ordinate system with origin on the axis of the cylinder, Ox parallel to the direction of U_0, Oz along the axis of the cylinder, and Oy at right angles to U_0 and to the cylinder axis. In the usual notation, the components of mean velocity at any point are U, V, 0, and the instantaneous velocity components are $U+u$, $V+v$, w. The instantaneous pressure fluctuation from the mean value is denoted by p.

Air density is denoted by ρ, and kinematic viscosity by ν.

All mean values are taken with respect to time at a fixed point.

THEORY OF WAKE FLOW

Using only the equations of motion of a viscous fluid without any further assumptions, comparatively few deductions may be made concerning the motion in a turbulent wake. Two distinct assumptions are always used to simplify the problem:

(*a*) that the transverse mean velocity component V is always small compared with the stream velocity U_0, equivalent to the usual approximations of boundary-layer theory;

(*b*) that the wake structure is similar at all distances down-stream from the cylinder, i.e. that the mean value of any quantity may be represented as a non-dimensional function of the co-ordinates, a local velocity which is a function of x only, and a local scale also a function of x only.

For example, the mean velocity is of the form

$$U_0 - U = U_1(x)\, f(y/l),$$

where

$$l = g(x).$$

There can be little argument over the accuracy of the first assumption, at least at positions not too close to the cylinder, but, by itself, it does not go very far. The equation of mean motion becomes

$$U_0\frac{\partial U}{\partial x} = -\frac{\partial \overline{uv}}{\partial y} + \nu\frac{\partial^2 U}{\partial y^2},$$

and the equation for the conservation of kinetic energy may be simplified to

$$\overline{W}/\rho + \tfrac{1}{2}U_0\frac{\partial}{\partial x}(\overline{u^2}+\overline{v^2}+\overline{w^2}) + \overline{uv}\frac{\partial U}{\partial y} + \frac{1}{2}\frac{\partial}{\partial y}(\overline{u^2v}+\overline{v^3}+\overline{vw^2}) + \frac{1}{\rho}\frac{\partial(\overline{pv})}{\partial y} = 0,$$

where the mean energy dissipation is

$$\overline{W} = \rho\nu\left[2\overline{\left(\frac{\partial u}{\partial x}\right)^2} + 2\overline{\left(\frac{\partial v}{\partial y}\right)^2} + 2\overline{\left(\frac{\partial w}{\partial z}\right)^2} + \overline{\left(\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\right)^2} + \overline{\left(\frac{\partial w}{\partial y}+\frac{\partial v}{\partial z}\right)^2} + \overline{\left(\frac{\partial u}{\partial z}+\frac{\partial w}{\partial x}\right)^2}\right].$$

Making use of the confirmed existence of local isotropy in the turbulent wake, this expression for the mean energy dissipation may be reduced to

$$\overline{W} = 15\rho\nu\overline{\left(\frac{\partial u}{\partial x}\right)^2}$$

(Townsend 1948*a*).

The second assumption depends on the notion that, after a sufficiently long time, the processes of the wake will settle down to a statistical equilibrium in which the velocity pattern is self-perpetuating, merely changing scale and intensity with lapse of time. It is worth noticing that a similar assumption has been successful in the decay of isotropic turbulence, but that very large eddies must be excluded from the similarity hypothesis, and it is not impossible that application of the similarity principle to the wake may be similarly restricted. In recent observations (Townsend 1947*a*), it has been shown that complete similarity does not exist in the wake if x/d is less than 1000, and, in the present work, no use has been made of the principle.

The hypothesis may still be useful for the reduction of results, and results derived from it by dimensional reasoning have approximate validity over not too wide ranges of x/d. These results may be summarized as follows: (*a*) the wake scale varies as $(x-x_0)^{\frac{1}{2}}$, and (*b*) the typical velocity varies as $(x-x_0)^{-\frac{1}{2}}$, where x_0 is the virtual origin of the wake.

Further progress may only be made by making additional assumptions about the physical nature of the turbulent movements. The most popular assumptions are those used in the various forms of the mixing length theory, which all suppose the scale of the turbulence to be small compared with the wake width, and the eddy diffusion time to be short compared with the time necessary for appreciable development of the wake. It will be shown that neither of these propositions is true, and consequently detailed confirmation of the theory is not possible, and the theories can only be applied to the deduction of integral properties such as mean velocity and temperature distributions. Briefly, the theory states that, if M is a transferable property, then the mean rate of transport of M is

$$\overline{mv} = -k\frac{\partial M}{\partial y},$$

where $k = v_1 l$ is the eddy diffusion coefficient, l is the 'mixing length', v_1 is the mixing velocity, and m is the instantaneous fluctuation of M from the local mean. Reasonable assumptions are then made regarding the variation of v_1 and l, usually relating v_1 and l by

$$v_1 = l\left|\frac{\partial U}{\partial y}\right|,$$

and then assuming that l is a constant over the wake. The time taken for an elementary diffusion process is then

$$\frac{l}{v_1} = 1\bigg/\left|\frac{\partial U}{\partial y}\right|,$$

and, in a typical wake, has a minimum value of nearly x/U_0, i.e. an elementary diffusion process takes at least as long as the mean stream takes to travel from the cylinder to the point of observation. Substituting experimental values of turbulent velocity and scale for the values predicted by the mixing length theory hardly affects the result, and so the second of the basic assumptions of mixing length theory is completely unjustified. The mixing length l is commonly one-fifth of the half-width of the wake, and in this respect the theory is also at fault, though not nearly to the same extent.

It is also possible to assume k a constant, but predictions of the mean velocity distribution are not consistent with experiment near the edge of the wake, though the agreement over the central portion of the wake is reasonable. The use of a diffusion coefficient, constant or not, can only be justified if the scale of the turbulent motions responsible for diffusion is small compared with the wake width in order that local mean conditions may be adequately described by the gradient $\partial M/\partial y$.

Hu (1944) has attempted to apply the statistical theory of turbulence to the problem, employing the methods of Chou (1945), who by making suitable assumptions has obtained explicit solutions. Many of the assumptions are perfectly reasonable, and, indeed, several have been confirmed experimentally in this laboratory, but, as will be seen below, the assumption of a constant energy-diffusion coefficient is almost certainly incorrect. It is not surprising then that predictions of the distribution of turbulent intensity in the wake have been inaccurate, and at present it seems unlikely that the statistical theory unaided can solve the problem.

EXPERIMENTAL DETAILS

All the measurements described were made in the wake of a circular cylinder, 0·953 cm. diameter, placed in an air-stream of velocity 1280 cm.sec.$^{-1}$. At the average temperature of the experiments, the cylinder Reynolds number was $U_0 d/\nu = 8130$. The air-stream was that of the small wind-tunnel in the Cavendish Laboratory, in which the free tunnel turbulence is

$$\sqrt{\overline{u^2}}/U_0 = 0{\cdot}0006, \quad \sqrt{\overline{v^2}}/U_0 = \sqrt{\overline{w^2}}/U_0 = 0{\cdot}0015 \quad \text{(Batchelor \& Townsend 1947)}.$$

Mean velocities in the wake were measured with a small total-head tube, 1 mm. outside diameter, and the turbulence levels were sufficiently low that no correction to the total-head reading was necessary.

The basic measuring instrument for turbulence is the hot-wire anemometer, used in the single-wire form for the observation of down-stream fluctuations, and in both the V- and X-wire forms for the observation of cross-stream fluctuations. The X-wire is more compact and less liable to systematic error, but results obtained with it were nearly identical with results from a V-form wire. By means of a simple double Wheatstone bridge, it is possible to obtain simultaneously from a V- or X-wire independent electric signals proportional to the instantaneous values of the down- and cross-stream fluctuations, which are necessary for the measurements of the velocity products $\overline{uv}$ and $\overline{u^2v}$. Calibration of such wires was carried out by placing them in the uniform mean flow behind a biplane grid, where the turbulent motion is known to possess ordinary isotropy. Not only can the sensitivity of a wire be measured in this way, but also it is possible to determine the directional characteristics, information necessary for the correct interpretation of the results, particularly for measurements of the velocity product $\overline{uv}$.

The results were obtained from three traverses of the wake, at distances of respectively 80, 120 and 160 diameters down-stream from the cylinder. Measurements were made of the following quantities:

(*a*) Turbulent intensities: $\overline{u^2}$, $\overline{v^2}$, $\overline{w^2}$.

(*b*) Mean squares of the velocity derivatives:

$$\overline{\left(\frac{\partial u}{\partial x}\right)^2}, \quad \overline{\left(\frac{\partial v}{\partial x}\right)^2}, \quad \overline{\left(\frac{\partial w}{\partial x}\right)^2}.$$

(*c*) Double-velocity product: $\overline{uv}$.

(*d*) Triple velocity products:

$$\overline{u^3}, \quad \overline{u^2v}, \quad \overline{v^3}, \quad \overline{vw^2}.$$

(*e*) Skewness factor: $$S_0 = \overline{\left(\frac{\partial u}{\partial x}\right)^3} \Big/ \left[\overline{\left(\frac{\partial u}{\partial x}\right)^2}\right]^{\frac{3}{2}}.$$

(*f*) Flattening factors:

$$\overline{u^4}/(\overline{u^2})^2, \quad \overline{v^4}/(\overline{v^2})^2, \quad \overline{w^4}/(\overline{w^2})^2,$$

$$\overline{\left(\frac{\partial u}{\partial x}\right)^4} \Big/ \left[\overline{\left(\frac{\partial u}{\partial x}\right)^2}\right]^2, \quad \overline{\left(\frac{\partial v}{\partial x}\right)^4} \Big/ \left[\overline{\left(\frac{\partial v}{\partial x}\right)^2}\right]^2, \quad \overline{\left(\frac{\partial w}{\partial x}\right)^4} \Big/ \left[\overline{\left(\frac{\partial w}{\partial x}\right)^2}\right]^2.$$

(*g*) Downstream scales: $$L_u = U\int_0^\infty \frac{\overline{u(t)\,u(t+\tau)}}{\overline{u^2}}\,d\tau,$$

$$L_v = U\int_0^\infty \frac{\overline{v(t)\,v(t+\tau)}}{\overline{v^2}}\,d\tau,$$

$$L_w = U\int_0^\infty \frac{\overline{w(t)\,w(t+\tau)}}{\overline{w^2}}\,d\tau,$$

where $u(t)$, $v(t)$, $w(t)$ are velocity components at a fixed point at time t.

To give a detailed account of the procedures for measuring the various quantities and the necessary electrical circuits would occupy considerable space, and the fundamental circuits have already been described (Townsend 1947*b*). The push-pull capacitance-compensated amplifiers are identical with those previously used, but the circuit for the measurement of triple-velocity products differs from the earlier version in detail but not in essentials. Briefly, the original squaring circuit is used to form an electric signal proportional to the instantaneous square of an input e_1, and to this 'squared' signal may be added the second input e_2. Since the amplifier does not pass a steady signal, the resultant is proportional to $(e_1^2-\overline{e_1^2}+\alpha e_2)$, where α is an instrumental constant. The mean square of this signal is then measured with a vacuum thermo-junction and a millivoltmeter, giving a deflexion proportional to

$$\overline{[e_1^2-\overline{e_1^2}+\alpha e_2]^2} = \overline{e_1^4}-(\overline{e_1^2})^2+2\alpha\overline{e_1^2e_2}+\alpha^2\overline{e_2^2}.$$

Then by suitable arrangements and reversals of the inputs, it is possible to measure the quantities $$\overline{e_1^4}, \quad \overline{e_1^2e_2}, \quad \overline{e_1^3}.$$

To calibrate the circuits, a standard wave-form is necessary, and both a pure sine-wave and the composite form ($\sqrt{2}\cos pt+\cos 2pt$) were used to provide independent calibrations.

The accuracy of the observations varies considerably, from 1 to 2 % in simple intensity measurements to nearer 5 % in some measurements of triple products In general, the results are reduced to non-dimensional form by use of the mean stream velocity U_0 and the cylinder diameter d.

From the measurements, it is obvious that the wake structure has not reached the dynamic equilibrium required by the similarity hypothesis, i.e. the dimensionless functions describing the wake motion are not independent of x. In spite of this, it is thought that the results at $x/d = 160$ are fairly typical of conditions in the region of dynamical similarity that is presumably being approached, and therefore these results would be more appropriate for comparison with theoretical predictions. The similarity hypothesis has not been used in the analysis of the results.

INTERMITTENCY OF TURBULENT FLOW IN THE WAKE

Even the most casual observation of oscillographic records of the turbulent fluctuations shows that, except close to the wake centre, the flow at any point is only fully turbulent for a fraction, γ, of the total time of observation, and that γ decreases with distance from the wake centre. (A similar result was obtained by Corrsin (1943) in a study of a turbulent jet.) γ is most readily calculated from measurements of the flattening factors, by assuming that the flattening factors taken only during periods of turbulent flow are identical with those observed in wholly turbulent flow, in particular in isotropic turbulence where experiment has shown that

$$\overline{\left(\frac{\partial u}{\partial x}\right)^4}\Big/\left[\overline{\left(\frac{\partial u}{\partial x}\right)^2}\right]^2 = 3{\cdot}50, \quad \overline{\left(\frac{\partial v}{\partial x}\right)^4}\Big/\left[\overline{\left(\frac{\partial v}{\partial x}\right)^2}\right]^2 = \overline{\left(\frac{\partial w}{\partial x}\right)^4}\Big/\left[\overline{\left(\frac{\partial w}{\partial x}\right)^2}\right]^2 = 3{\cdot}00,$$

$$\overline{u^4}/(\overline{u^2})^2 = \overline{v^4}/(\overline{v^2})^2 = \overline{w^4}/(\overline{w^2})^2 = 3{\cdot}00.$$

Then $$\gamma = 3{\cdot}5\left[\overline{\left(\frac{\partial u}{\partial x}\right)^2}\right]^2\Big/\overline{\left(\frac{\partial u}{\partial x}\right)^4} = 3{\cdot}0\left[\overline{\left(\frac{\partial v}{\partial x}\right)^2}\right]^2\Big/\overline{\left(\frac{\partial v}{\partial x}\right)^4} = 3{\cdot}0\left[\overline{\left(\frac{\partial w}{\partial x}\right)^2}\right]^2\Big/\overline{\left(\frac{\partial w}{\partial x}\right)^4}$$

$$= 3{\cdot}0(\overline{u^2})^2/\overline{u^4} = 3{\cdot}0(\overline{v^2})^2/\overline{v^4} = 3{\cdot}0(\overline{w^2})^2/\overline{w^4}.$$

The assumption is justified by the consistency of values of γ measured in this way (figure 1). It is also a prediction of the theory of local isotropy, valid when statistical equilibrium for small eddies exists, that the flattening factors of $\partial u/\partial x$, $\partial v/\partial x$, $\partial w/\partial x$ should be absolute constants, and consequently that the first three expressions for γ should give the same result. It may be noted that, at the extreme edge of the wake, values of γ calculated from the flattening factors of u, v and w are systematically larger than the others. This is due to the dominant role of large eddies in this region, and the result is perfectly consistent with the postulated structure.

It is concluded that there are regions of fully developed turbulent flow in which statistical equilibrium for small eddies and local isotropy have been established, separated by regions of nearly purely laminar motion. It is reasonable to suppose that the intermittency of turbulent flow is due to jets or outflows of turbulent fluid from the inner core of fully turbulent fluid, and that the shape of the outflows and their statistical distribution in space together determine the variation of γ across the wake. That is, the diminution of γ toward the edge of the wake is due in part to the mean width of the jets being less there, and in part to the lower probability that a jet should penetrate so far from the central core. A hot-wire placed outside the central core then only records turbulence while one of these jets is being swept past it. In figure 2, an attempt is made to represent schematically an instantaneous section of the wake, and plate 24*b* in Goldstein's *Modern developments in fluid dynamics*, an instantaneous photograph of a smoke jet, resembles closely the type of motion that is envisaged. Development of existing jets and initiation of new ones must proceed continuously in order that the observed spread of the wake should be possible. It must be emphasized that the intermittency is not confined to the extreme edge of the wake, but is appreciable even closer to the wake centre than the region of maximum shear, and that the only reasonable conclusion is that intermittency is not a secondary phenomenon, but plays an essential part in the motion and development of the wake.

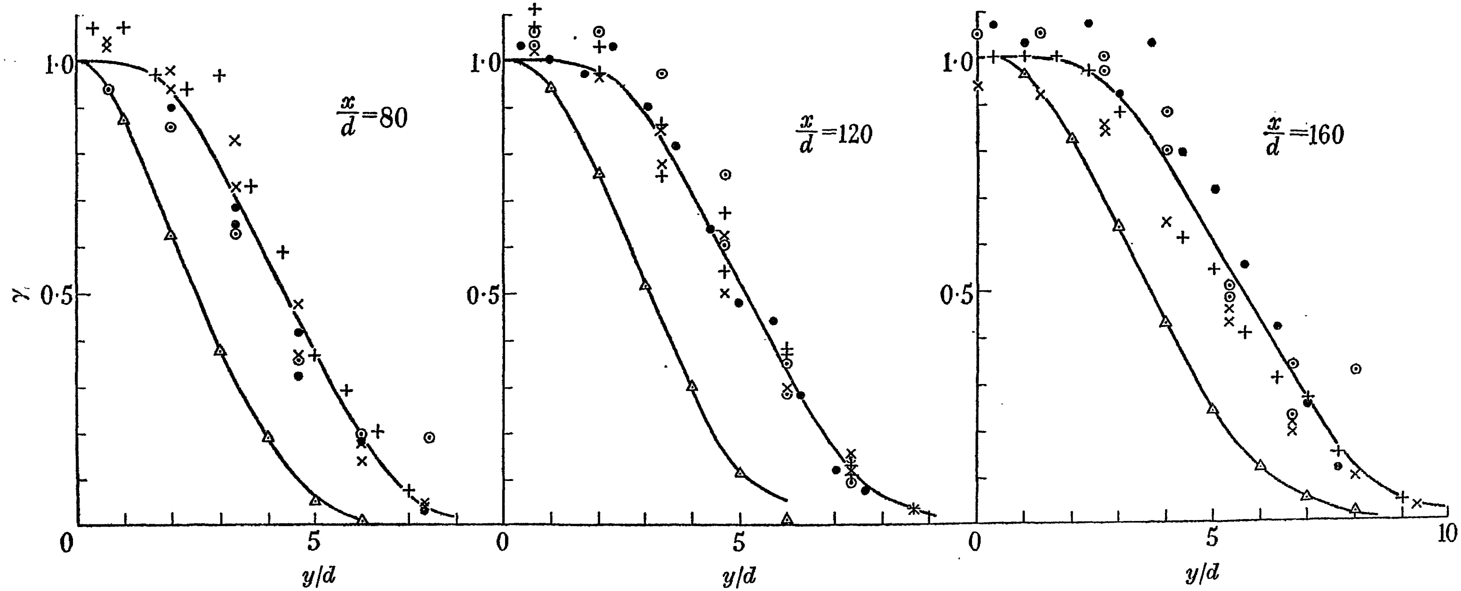


FIGURE 1. Intermittency factors in cylinder wake.

● $\gamma = 3{\cdot}5\left[\overline{\left(\frac{\partial u}{\partial x}\right)^2}\right]^2 \Big/ \overline{\left(\frac{\partial u}{\partial x}\right)^4}$.

× $\gamma = 3{\cdot}0\left[\overline{\left(\frac{\partial w}{\partial x}\right)^2}\right]^2 \Big/ \overline{\left(\frac{\partial w}{\partial x}\right)^4}$.

+ $\gamma = 3{\cdot}0\left[\overline{\left(\frac{\partial v}{\partial x}\right)^2}\right]^2 \Big/ \overline{\left(\frac{\partial v}{\partial x}\right)^4}$.

⊙ $\gamma = 3{\cdot}0(\overline{u^2})^2/\overline{u^4}$.

△ $(U_0^2 - U^2)/(U_0^2 - U_{\text{min.}}^2)$.

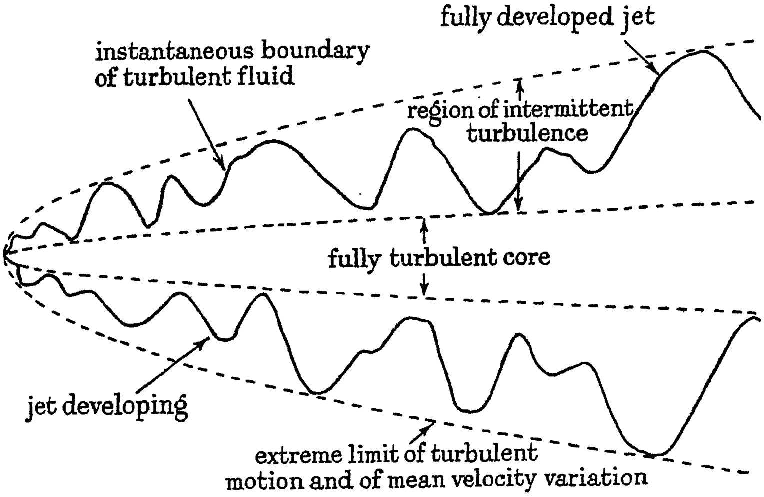

fully developed jet
instantaneous boundary of turbulent fluid
region of intermittent turbulence
fully turbulent core
jet developing
extreme limit of turbulent motion and of mean velocity variation

FIGURE 2. Section of hypothetical wake structure.

STRUCTURE OF THE WAKE

Let us now consider the experimental results in turn, and use them to derive information about the detailed properties of these jets of turbulent fluid. In the first place, the velocity product $\overline{uv}$, representing the Reynolds shear stress, has been measured, and, with the observed distribution of mean velocity across the wake, the effective eddy viscosity ϵ and the experimental mixing length l may be calculated, using the definitions

$$\overline{uv} = -\epsilon\frac{\partial U}{\partial y}, \quad \epsilon = l\sqrt{\overline{v^2}}.$$

Experimentally, l is found to be fairly small, approximately 0·07 of the half-width of the mean velocity wake (figure 3), and does not vary greatly over the width of the wake. The small size of l is interpreted as evidence that momentum transfer in the wake is carried out by comparatively small eddies. More significantly, ϵ/γ is not far from constant over the greater part of the wake (figure 4), but this will be discussed later.

The product $\overline{uv}$ may be regarded as the rate of transport of momentum (per unit mass), and similarly the rate of transport of turbulent energy is

$$\tfrac{1}{2}(\overline{u^2v}+\overline{v^3}+\overline{vw^2}),$$

and, in principle, it is possible to calculate an energy diffusion coefficient δ, analogous with ϵ, by use of the defining equation

$$\overline{u^2v}+\overline{v^3}+\overline{vw^2} = -\delta\frac{\partial}{\partial y}(\overline{u^2}+\overline{v^2}+\overline{w^2}).$$

When this is attempted (figure 5), no simple behaviour is found either for δ, or for the corresponding mixing length. Negative values occur near the wake centre, and, even where the turbulence gradient is fairly uniform, δ remains large compared with ϵ,

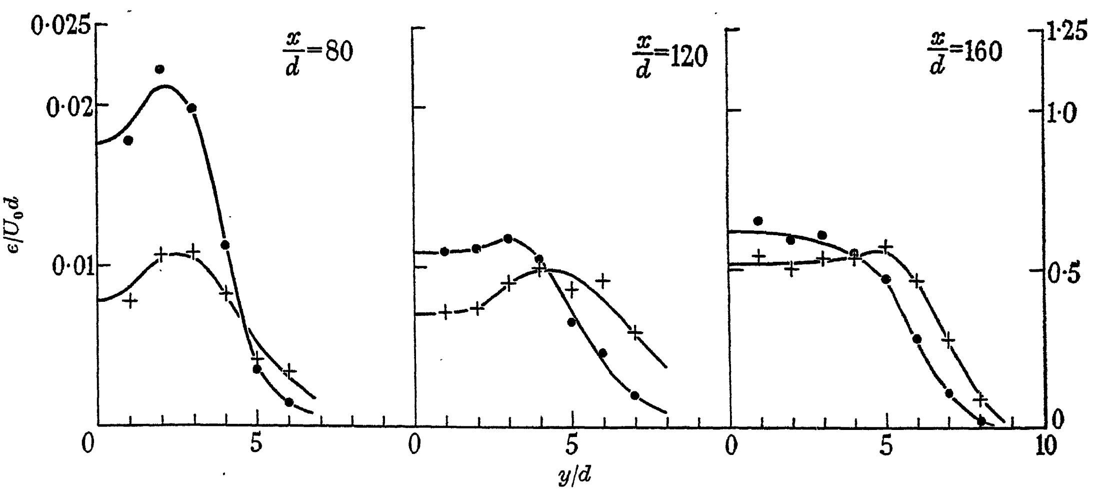


Figure 3. Measurements of effective viscosity and mixing length. ($U_0 = 1280$ cm.sec.$^{-1}$, $d = 0·953$ cm.) ● = $\epsilon/U_0 d$. + = l/d.

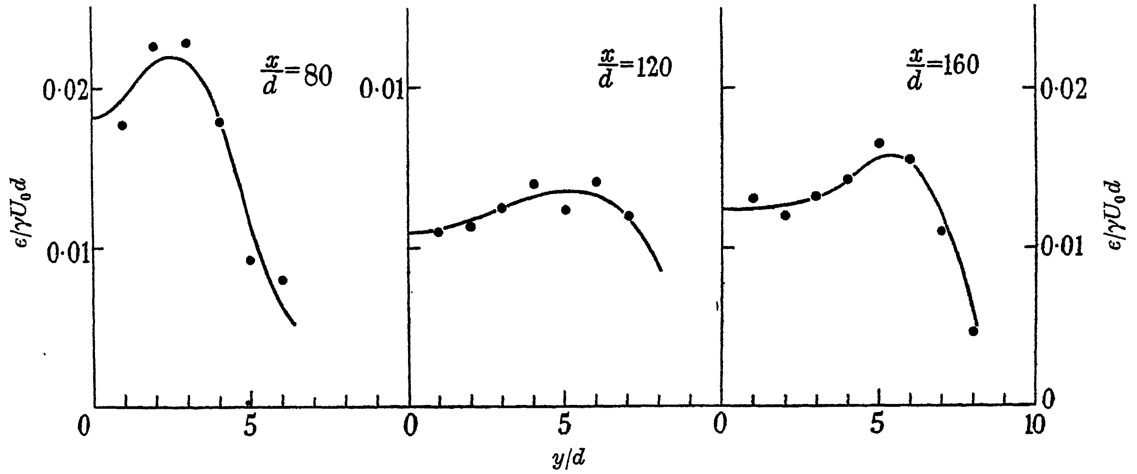


Figure 4. Measurements of jet effective viscosity. ($U_0 = 1280$ cm.sec.$^{-1}$, $d = 0·953$ cm.)

and decreases rapidly with distance from the wake centre. It must be concluded that the use of a diffusion coefficient to describe the transport of turbulent energy is not justified, and that energy diffusion is a process independent of momentum diffusion.

To remove this difficulty, it is not sufficient to consider the effects of intermittency. If the intermittency factor is known, then the mean intensity in the turbulent regions is

$$I_j = \frac{\overline{u^2}+\overline{v^2}+\overline{w^2}}{\gamma},$$

and I_j is found to vary only slightly over the greater part of the wake (figure 6). So a considerable transport of energy is found in the almost complete absence of a real intensity gradient, and it is difficult to see how energy flow can take place by turbulent

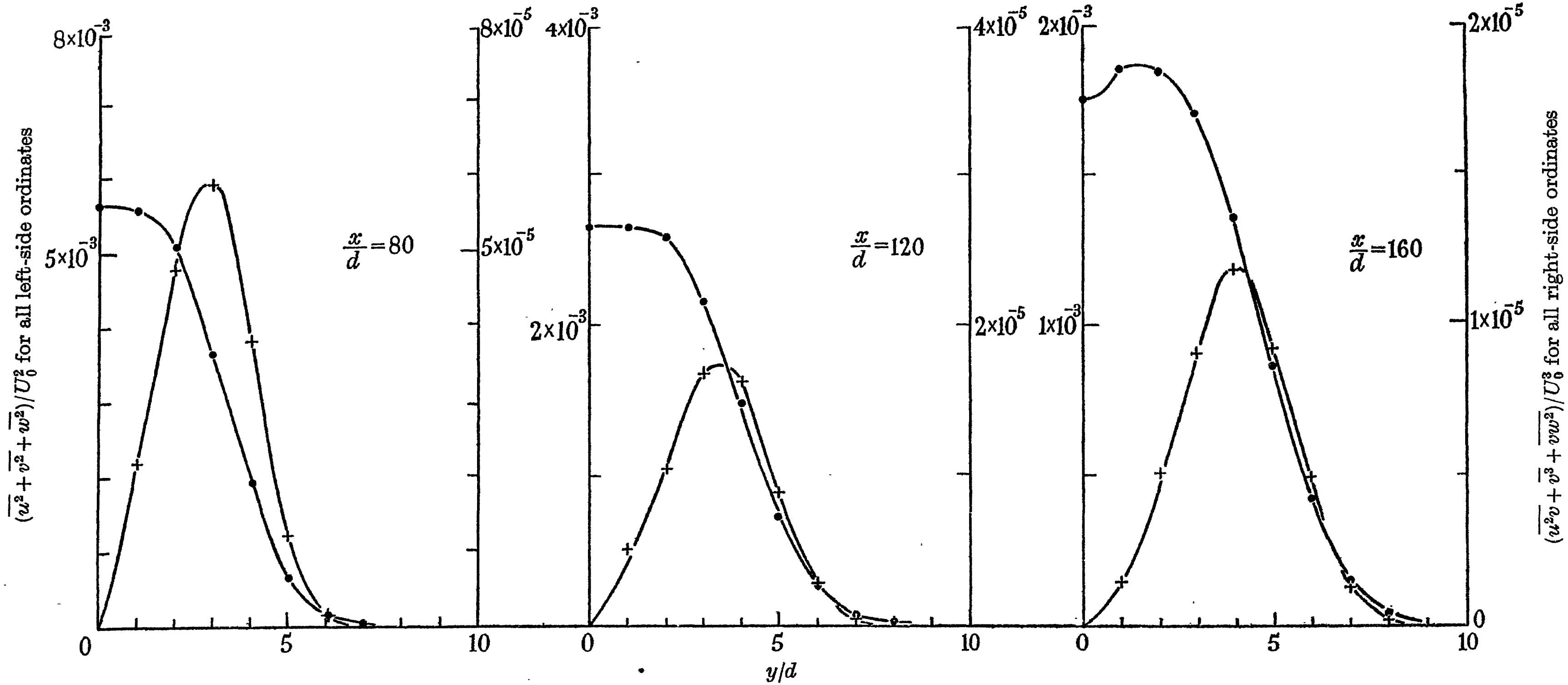


FIGURE 5. Measurements of turbulent energy and of turbulent energy diffusion.

($U_0 = 1280$ cm.sec.$^{-1}$, $d = 0{\cdot}953$ cm.) $\bullet = (\overline{u^2}+\overline{v^2}+\overline{w^2})/U_0^2$. $+ = (\overline{u^2v}+\overline{v^3}+\overline{vw^2})/U_0^3$.

movements inside the jets. For the transport mechanism, there is only left the bulk movement of the jets, which is naturally outwards and away from the wake centre. The compensating inflow will consist of non-turbulent fluid transporting no turbulent energy. Consequently, the flow of energy is not dependent on the local intensity gradient (if any), but only on the mean jet velocity and the jet turbulent intensity, which in turn are determined by conditions in the turbulent core.

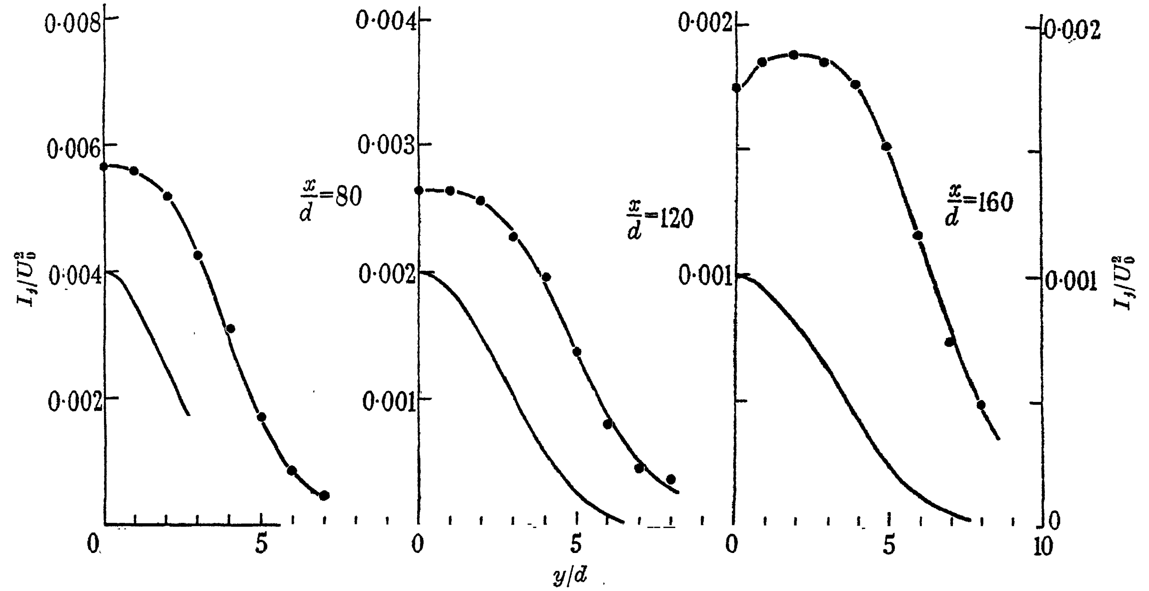


FIGURE 6. Measurements of $I_j = (\overline{u^2}+\overline{v^2}+\overline{w^2})/\gamma$ across the wake.
(N.B. Lower curve represents variation of mean velocity on an arbitrary scale.)

Algebraically, if the mean jet velocity is $\mathscr{V}$, then the mean rate of flow of turbulent intensity is

$$\mathscr{V}(\overline{u^2}+\overline{v^2}+\overline{w^2}) = \overline{u^2v}+\overline{v^3}+\overline{vw^2}.$$

$\mathscr{V}$ may be calculated and the results are given in figure 7. It may be noted that diffusion is relatively more violent at $x/d = 80$ where similarity is not even approximately valid, but the mean jet velocities are all considerably less than half the local turbulent velocities, and this fits in well with the notion that the jets themselves carry little energy except in the form of turbulence. Moving away from the wake centre, the mean jet velocity increases about parabolically to a maximum quite close to the edge of the wake, and then decreases sharply to zero.

Considering again the transport of momentum, two mechanisms are possible in this structure, momentum transport by bulk movement of the jets, or by turbulent motion inside the jets. This distinction is drawn since there must be at least an order of magnitude difference in scale between these motions that the observed sharp distinction between turbulent and non-turbulent fluid may exist. In view of the difference in character of the distributions and transport rates of turbulent intensity and mean flow momentum, it is tempting to propose that energy transport is solely due to jet movement (as is probably true), and that momentum transport is only carried on by turbulent motion in the jets. In support of this hypothesis, it may be urged that ϵ/γ, which may be called the jet eddy viscosity, is nearly constant throughout the wake, but a similar result is possible, even if momentum transport is due to the mean jet motion. Better evidence is provided by measurements of the longitudinal scale L_u (figure 8). If the slow jet movements cause momentum diffusion,

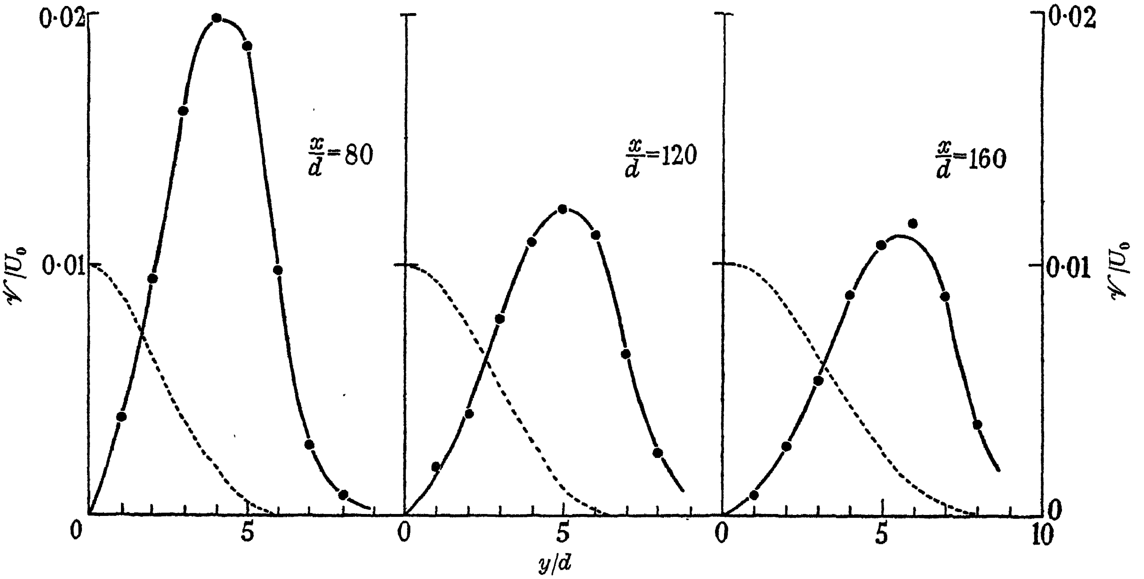


FIGURE 7. Measurements of mean jet velocity $\mathcal{V}$.
(N.B. Dashed curve represents variation of mean velocity on an arbitrary scale.)

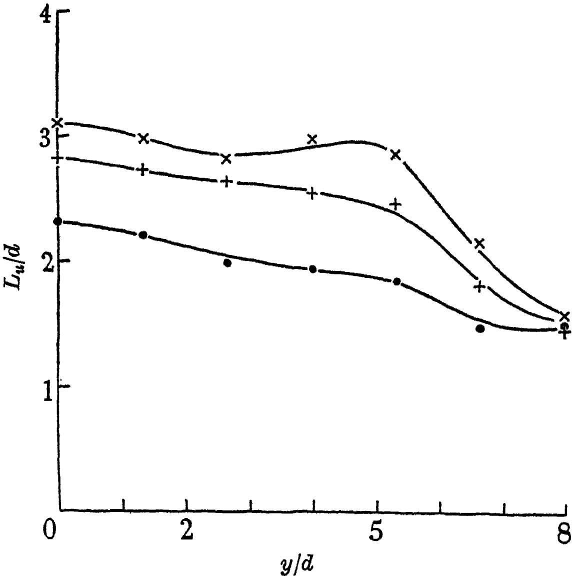


FIGURE 8. Measurements of longitudinal scale. • $x/d = 80$, + $x/d = 120$, × $x/d = 160$.

then the mixing processes occur rather infrequently, though perhaps in large amounts and large scale variations of downstream velocity would be expected. Experimentally, L_u is found either to decrease slightly or to remain constant in the region of shear, and it is concluded that the mixing processes are frequent, and therefore depend on the turbulent motion inside the jets rather than on the mean jet movements.

The structure under discussion consists of a fully turbulent core which emits jets or billows of turbulent fluid into the undisturbed fluid around it. The jets may be large eddies, carrying with them small eddies in dynamic equilibrium with each other, and observations show their boundaries to be fairly sharp. They are probably caused by local fluctuations either of the local Reynolds stresses or of the pressure, and are both large and slow-moving. The maximum size of the eddies inside the jets is at least an order of magnitude less than the jet width, and they are responsible for the eddy viscosity, i.e. they are the dominant factor in the transfer of momentum. The mean jet motion must also transport momentum in some quantity, but it seems that this transport is insufficient to affect the result appreciably. Inside the jets, there is no significant intensity gradient and the transport of kinetic energy is determined by the mean jet motion only, and, since this motion is connected with the largest eddies of the central core, it is probably slow but of considerable range, sufficient to transfer turbulent fluid from the central core to the outermost part of the wake.

THE ENERGY BALANCE AND PRESSURE FLOW IN THE WAKE

At any point in the wake, the mean turbulent intensity (averaged over a sufficiently long time interval) is constant, and the gain of turbulent energy by convection, diffusion, work against Reynolds shear stress and pressure flow equals the loss by viscous dissipation. Making the usual boundary layer approximations and assuming further that local isotropy exists, the equation for the conservation of energy reduces to

$$15\nu\left(\frac{\partial u}{\partial x}\right)^2+\tfrac{1}{2}U_0\frac{\partial}{\partial x}(\overline{u^2}+\overline{v^2}+\overline{w^2})+\overline{uv}\frac{\partial U}{\partial y}+\frac{1}{2}\frac{\partial}{\partial y}(\overline{u^2v}+\overline{v^3}+\overline{vw^2})+\frac{1}{\rho}\frac{\partial\overline{pv}}{\partial y}=0.$$

In the course of these experiments, all the terms excepting the pressure term have been measured directly, and the pressure term has then been obtained by difference (figure 9). No particular comment on the general behaviour of the directly measured terms is necessary, but it is interesting that in the region of shear the contribution of the pressure term to the energy balance is comparable with that of the diffusion term. Outside the mean velocity wake, it appears that spread of turbulent motion is almost entirely due to pressure flow, and, in this region, the equation may be simplified further to

$$\tfrac{1}{2}U_0\frac{\partial}{\partial x}(\overline{u^2}+\overline{v^2}+\overline{w^2})+\frac{1}{\rho}\frac{\partial\overline{pv}}{\partial y}=0.$$

The negligible part played by diffusion and dissipation in this region suggests that the motion approximates closely to a potential flow.

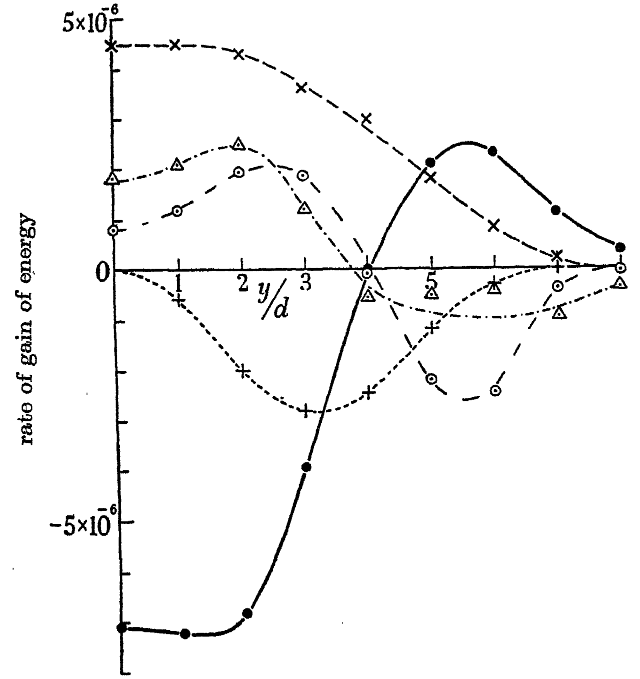


Figure 9. Energy balance in the wake.

● $\dfrac{1}{2}\dfrac{\partial}{\partial x/d}\left(\dfrac{\overline{u^2}+\overline{v^2}+\overline{w^2}}{U_0^2}\right)$ convection. ⊙ $\dfrac{1}{2}\dfrac{\partial}{\partial y/d}\left(\dfrac{\overline{u^2v}+\overline{v^3}+\overline{vw^2}}{U_0^2}\right)$ diffusion. + $\dfrac{\overline{uv}}{U_0^3}\dfrac{\partial U}{\partial y/d}$ shear.

× $15\dfrac{\nu}{U_0 d}\left[\dfrac{\partial\left(\dfrac{u}{U_0}\right)}{\partial x/d}\right]^2$ viscous dissipation. △ $\dfrac{1}{\rho U_0^3}\dfrac{\partial\overline{pv}}{\partial y/d}$ pressure flow.

MECHANISM OF THE WAKE

A structure of the cylinder wake has been proposed, partly by induction from the experimental results, but also from pure guess-work, but little has been said about the forces necessary to sustain such a structure. The most striking feature of the wake is the production and motion of the billowing outflows of turbulent fluid, which for convenience have been termed jets, and two possible explanations have been mentioned: (*a*) that they represent the 'over-shoot' of turbulent movements inside the core, a statement equivalent to saying that fluctuations in the turbulent Reynolds stresses cause the outflows; and (*b*) that local pressure fluctuations cause acceleration of turbulent fluid away from the wake centre, so producing a jet.

If the first hypothesis is true, the jet commences by moving rapidly and then probably slows down, but second hypothesis demands that during its formation the jet is accelerated, and only begins to slow down or to turn around when it is nearly fully developed. Measurements of $\mathscr{V}$, the mean jet velocity (figure 7), strongly suggest that pressure flow does produce the jets, and the indirect determination of the pressure-velocity correlation confirms the existence in the wake of pressure fluctuations of appropriate magnitude. The eventual stopping of the jet is probably due to adverse pressure gradients, possibly associated with the increase of virtual mass of the jet as it emerges from the core. It may be noted that dissipation and diffusion at the extreme edge of the wake are negligible, suggesting that there the motion approximates to a potential flow, and this would be characteristic of slowing by increase of virtual mass.

The maintenance and development of the mean flow pattern has been supposed to be due to the shear forces which occur only in the regions of turbulent motion, but the intervening portions of non-turbulent fluid must be similarly accelerated. It is assumed that this is done by longitudinal pressure gradients, but, in the absence of a practicable direct method of measuring pressure fluctuations, it is not possible to investigate the matter in detail.

A much more important matter is the increase of volume of turbulent fluid, which must occur continuously and at a considerable rate. The results show that, due to intermittency, the apparent volume is much greater than the real volume, but still the increase exists and must be fitted into the scheme. The increase occurs at the boundaries of the turbulent regions by encroachment on the undisturbed fluid by diffusive motions of the smaller eddies comprising the jet turbulent motion. Previously, some pains have been taken to minimize this effect in comparison with mean jet motion, due to the mean jet motion, the surface area of the boundaries is large and even slow diffusion should suffice to produce the necessary increase of turbulent volume. It may be put that the turbulent diffusion occurs in two stages, the mean jet movement doing the large scale diffusion and the jet turbulence performing small scale diffusion of the resulting pattern; that is, the jet motion acts to increase the surface area of the turbulent fluid, and so allow a comparatively slow linear diffusion at the bounding surfaces to achieve a large rate of volume increase.

The reason for the occurrence of such a two-stage turbulent pattern is not entirely clear, but it appears that the scale of the mean flow is intermediate between the scale of the mean jet motion and the scale of the jet turbulent motion. This suggests that the presence of the stable mean flow system with energy larger than the total of turbulent energy leads to production of some extremely large eddies as well as the more usual smaller ones. Thinking in terms of the three-dimensional Fourier series representation of the velocity field, the effect of the non-linear inertia terms in the equations of motion is to intermodulate existing Fourier components introducing new sum and difference wave-numbers. Considering the mean flow system as a narrow band of Fourier components of high specific energy, the difference wave-numbers all represent large eddies which may well be the jets, while the summation wave-numbers may represent the jet turbulence. In any event, it is expected that this type of turbulent flow will be found in any shear flow unrestricted in extent, i.e. in jets and wakes, and in boundary layers on the free side.

FURTHER DEVELOPMENTS

The mean jet velocity $\mathscr{V}$ should be equally valid for the diffusion of any transferable quantity, in particular for the diffusion of heat. That is, the rate of heat transfer per unit heat capacity should be

$$\overline{\theta v} = \mathscr{V}(T - T_a),$$

where T is the mean local temperature, $T+\theta$ is the instantaneous local temperature, and T_a is the ambient temperature. An independent measurement of $\mathscr{V}$ in this way would be a valuable check on the theory, and will be undertaken shortly.

The jet structure can only be surmised at present, and many interesting questions remain to be answered. It is possible that useful information may be obtained from measurements of quadruple velocity products of the form $\overline{u_1^2 u_2^2}$, which are technically possible. In this way, more exact information about jet widths and rates of development could be collected, and a clearer understanding of the details of the motion attained.

REFERENCES

Batchelor, G. K. 1947 *Proc. Camb. Phil. Soc.* **43**, 533–559.
Batchelor, G. K. 1948 *Quart. Appl. Math.* **6**, 97–116.
Batchelor, G. K. & Townsend, A. A. 1947 *Proc. Roy. Soc.* A, **190**, 534–550.
Batchelor, G. K. & Townsend, A. A. 1948 *Proc. Roy. Soc.* A, **193**, 539–558.
Chou, P. Y. 1945 *Quart. Appl. Math.* **3**, 38–54.
Corrsin, S. 1943 *Nat. Adv. Ctee. Aero. War. Rep.* W-94.
Heisenberg, W. 1948 *Z. Phys.* **124**, 628–657
Hu, N. 1944 *Chinese J. Phys.* **5**, 30.
Kolmogoroff, A. N. 1941 *C.R. Acad. Sci. U.R.S.S.* **30**, 301.
Kolmogoroff, A. N. 1941 *C.R. Acad. Sci. U.R.S.S.* **32**, 16.
Onsager, L. 1945 *Phys. Rev.* **68**, 286 (Abstract only).
Townsend, A. A. 1947*a* *Proc. Roy. Soc.* A, **190**, 551–561.
Townsend, A. A. 1947*b* *Proc. Camb. Phil. Soc.* **43**, 560–570.
Townsend, A. A. 1948*a* *Austr. J. Sci. Res.* **1**, 161–174,
Townsend, A. A. 1948*b* *Proc. Camb. Phil. Soc.* **44**, 560–565.
v. Weizsacker, C. F. 1948 *Z. Phys.* **124**, 614–627.

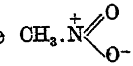
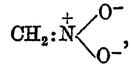
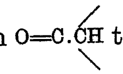
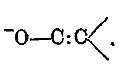
The catalyzed dehydration of acetaldehyde hydrate, and the effect of structure on the velocity of protolytic reactions

By R. P. Bell, F.R.S., and W. C. E. Higginson
Physical Chemistry Laboratory, University of Oxford

(*Received* 11 *November* 1948)

The reaction $CH_3CH(OH)_2 \rightarrow CH_3CHO + H_2O$ has been studied kinetically in aqueous acetone using a dilatometric method. It was found to be a first-order reaction, catalyzed by uncharged acids, cation acids, uncharged bases and anion bases. Quantitative measurements were made on the catalytic effect of 63 acids. The results for 32 carboxylic acids and 15 phenols conform approximately to a single relation between catalytic power and dissociation constant in water, but most of the 16 acids of other types show considerable positive or negative deviations from this relation. These deviations are related to the structures of the catalysts, and it is shown that, for acids of equal strength, the velocity of protolytic change is decreased by a mesomeric charge-shift in the anion, and is increased if mesomerism in the acid is lost on ionization. This principle is related to the occurrence of pseudo-acids and bases, and is interpreted in terms of molecular potential energy curves.

Introduction

Protolytic reactions between acids and bases normally take place so rapidly that their velocities cannot be measured, even by special experimental devices. There are, however, two classes of reaction in which the velocity is low enough to be measured. The first class concerns pseudo-acids and pseudo-bases, whose reactions with ordinary acid-base systems are slow enough to be studied by straightforward methods. This behaviour is always associated with a change of electronic structure in the pseudo-acid or base (Bell 1947): for example, nitromethane  $CH_3.\overset{+}{N}(=O)O^-$ reacts slowly with hydroxyl ions losing a proton to form the ion  $CH_2{:}\overset{+}{N}(O^-)O^-$, in which the negative charge is far removed from the atom to which the proton was attached. The second class of slow protolytic reaction is met with in general acid-base catalysis, which always involves the slow exchange of a proton between the substrate and the catalyst (Bell 1941, chap. vii). The substrate is normally such a weak acid or base that the extent of its reaction with the catalyst is too small to be detected by ordinary methods, and the possibility of studying the protolytic reaction depends on some further phenomenon, associated with the primary reaction, such as racemization, isotope exchange, decomposition, or rapid reaction with some other species. Here also the slow reaction is usually associated with a change in electronic structure in the substrate, which can be classed as a very weak pseudo-acid or base: for example, the base-catalyzed prototropy of ketones and similar substances involves the loss of a proton by the system  $O{=}C.CH\langle$ to give the ion  $^-O{-}C{:}C\langle$.

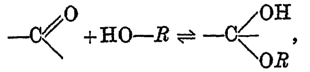
It seems probable that pseudo-acid-base behaviour in the sense described above represents an extreme case of a general relation between electronic structure and rate of proton transfer. Modern theories and experimental investigations of molecular structure indicate that a large proportion of acids and bases undergo some degree of electronic re-arrangement when losing or gaining a proton, which should give rise to differences in the velocities of their protolytic reactions (Bell 1940). It is unlikely that such differences could be detected by investigating ordinary acid-base reactions, since even in extreme cases like the nitro-paraffins the velocities of neutralization are only just low enough to be studied quantitatively at ordinary temperatures. More convenient velocities can be obtained if one of the reactants is throughout a pseudo-acid or base, i.e. in a catalyzed reaction where the substrate remains the same and the acid-base systems to be investigated are used as catalysts.

The velocity of such a catalyzed reaction will of course depend on the acid-base strength of the catalyst, even if structural effects are absent. It is now well established that for a series of structurally similar catalysts (e.g. carboxylic acids) this dependence is a quantitative one and can be expressed by a relation of the form $k = GK^{\alpha}_{A,B}$, where k is the catalytic constant, $K_{A,B}$ the acidic or basic dissociation constant of the catalyst, and G and α quantities depending only on the solvent, the temperature and the nature of the reaction, α being less than unity. This type of relation was first proposed by Brönsted, and is commonly known by his name. Most experimental work on the Brönsted relation has been concerned with series of very similar catalysts, and although the effect of charge on G and α has been demonstrated there has been no systematic attempt to investigate the effect of structure. A few isolated observations suggest that the effect may be a considerable one. Thus it has been found that in various reactions the picric acid molecule (Brönsted & Bell 1931; Bell 1934) and the ions of nitrourethane (Brönsted & Pedersen 1924) and nitramide (Tong & Olson 1941) have a much smaller catalytic effect than would be expected from the Brönsted relation valid for other catalysts of the same charge-type: in each case the addition or subtraction of a proton involves a considerable electronic shift in the catalyst.

The object of the work described here was to compare the catalytic effect of acids varying as widely as possible in structure. The reaction employed had not been previously studied kinetically, though reactions of this kind are believed to be of importance in many biological processes. Its value in the present context lies in the fact that it is sensitive to very weak acids and in the convenience and speed with which kinetic measurements can be made, thus making it possible to study a wide range of catalysts.

The hydration of acetaldehyde

Although the evidence for the existence of a solid hydrate $CH_3CHO.H_2O$ (analogous to chloral hydrate) is not very satisfactory (Colles 1906), there are strong reasons for supposing that in aqueous solution acetaldehyde is present largely as the hydrate $CH_3CH(OH)_2$. Ramsay & Young (1886) noted that acetaldehyde and water mix with considerable evolution of heat and diminution of volume, and their observations were confirmed and extended by Perkin (1887). Later quantitative work on the heat of mixing (Brown & Pickering 1897) and the density of aqueous solutions (Homfray 1905) showed that the degree of hydration varies with temperature and dilution, and indicated that a considerable proportion of aldehyde remains unhydrated unless a large excess of water is present. Investigations of ultra-violet absorption (Schou 1926; Herold & Wolf 1931) point to hydration, since the characteristic absorption of the carbonyl group is greatly diminished in dilute solutions of acetaldehyde in water and other hydroxylic solvents. Finally, Herbert & Lauder (1938) showed by means of the isotope O^{18} that there is a rapid exchange of oxygen between water and acetaldehyde, which they attribute to the reversible formation of the hydrate $CH_3CH(OH)_2$. The analogous exchange between acetone and water is a slow process even at 100° C (Cohn & Urey 1938), and there is also some evidence that the hydration equilibrium of acetaldehyde is not established instantaneously. Perkin (1887) found that on mixing acetaldehyde and water the immediate effect was a small absorption of heat, followed by a much greater heat evolution which was not complete after 24 min. He also noted that a change of temperature caused slow variations in the density of acetaldehyde-water mixtures, equilibrium being reached only after some hours. Herbert & Lauder (1938) also showed that the isotopic exchange takes place at a measurable rate. It is likely that this change will be accelerated by acids and bases. Brown & Pickering (1897) observed that the heat of hydration was evolved more rapidly in the presence of ammonia, and the analogous hydration equilibrium $CO_2 + H_2O \rightleftharpoons H_2CO_3$ is known to exhibit general base catalysis (Booth & Roughton 1938). Moreover, many acid-base catalyzed reactions involve the reversible addition of a hydroxy-compound to a carbonyl group according to the scheme



$$-C{\overset{\large O}{\diagup}} + HO{-}R \rightleftharpoons -C\!\begin{smallmatrix}\diagup OH\\ \diagdown OR\end{smallmatrix},$$

for example, the reversible formation of semi-acetals (Dieckmann 1916, 1917), the mutarotation of glucose (Lowry & Smith 1927; Brönsted & Guggenheim 1927), the reversible dimerization of hydroxy-aldehydes and ketones (Bell & Baughan 1937; Bell & Hirst 1939), the mutarotation of certain α-keto-esters (McKenzie & Mitchell 1929; McKenzie & Ritchie 1931, 1932) and exchange reactions between alcohols and esters or acetals (Schaefgen, Verhoek & Newman 1945; Alquier 1943). For the dehydration of acetaldehyde hydrate probable mechanisms would be as follows:

acid catalysis
$$CH_3CH(OH)_2 + HB \rightleftharpoons CH_3CH(OH)(\overset{+}{O}H_2) + B$$
$$\rightleftharpoons CH_3CHO + H_2O + HB$$

basic catalysis
$$CH_3CH(OH)_2 + B \rightleftharpoons CH_3CH(OH)O^- + BH^+$$
$$\rightleftharpoons CH_3CHO + H_2O + B.$$

METHOD OF MAKING KINETIC MEASUREMENTS

If an aqueous solution of acetaldehyde is added to a large amount of inert solvent, the hydrate dissociates at a measurable rate, producing a considerable volume increase. In the present work 2·5 ml. of an aqueous solution containing 14·5 % acetaldehyde by weight was added to 25·5 ml. of acetone containing the catalyst to be studied, and the mixture transferred at once to a simple dilatometer of about 20 ml. capacity having a capillary of 0·2 to 0·3 sq.mm. cross-section. The course of the reaction corresponded to 8 to 12 cm. movement in the capillary, and the position of the meniscus could be read to 0·05 mm. The dilatometer was filled by suction through a 1 mm. capillary attached to the other end of the bulb, and was sealed by forcing a small quantity of high-temperature tap-lubricant into the end of this capillary, thus avoiding the use of a tap. The acetone was evacuated for a few minutes before use to remove dissolved air, which would otherwise produce bubbles during the experiment. Measurements were made in a thermostat at 25 ± 0·003° C, corresponding to 0·02 mm. in the height of the meniscus. Reliable readings could be obtained within 5 min. of mixing, and about 40 readings were normally taken.

In presence of acid catalysts the reaction was strictly of the first order over at least 90 % of its course, though very long reaction times reveal a slow subsequent volume change corresponding to a reaction about 30 times slower than the main change. This may be due to the dehydration of other hydrates, such as $(CH_3CHOH)_2O$, since a solid hydrate $(CH_2Cl.CHOH)_2O$ has been isolated, and hydrates of this type are believed to exist in aqueous formaldehyde solutions (Walker 1944). First-order constants were obtained graphically from the first 90 % of the reaction by the method of Guggenheim (1926), and showed no systematic variation when the concentration of acetaldehyde or water was changed by a factor of eight. It might be suspected that the change observed was not the dehydration of the aldehyde hydrate, but involved a reaction with the acetone. However, this was disproved in two ways: first, closely similar results were obtained if dioxane was used as solvent in place of acetone, and secondly, no volume change was observed if acetaldehyde was added to a solution of water in acetone, or water added to a solution of acetaldehyde in acetone. It can be similarly shown that no appreciable reaction takes place under the conditions of our experiments between acetaldehyde and any of the catalysts used, though such reactions are known or would be expected to exist in some instances.

Some experiments were done with triethylamine and tri-*n*-butylamine as basic catalysts. When acetone is used as the solvent the initial increase in volume is followed by a slower decrease, which is probably due to a condensation reaction between acetaldehyde and acetone, since it is absent in dioxane solution. It is, however, possible to disentangle the two reactions and thus to determine the catalytic effect of the amine in the dehydration reaction. Experiments with salts of aliphatic and aromatic amines showed that the amine cations can act as acid catalysts just like the uncharged acids. Catalysis by amine molecules and cations has not yet been examined thoroughly, and it is mentioned here only to confirm the existence of general acid-base catalysis in its widest sense.

Experiments with other aldehydes showed that propionaldehyde and isobutyraldehyde form hydrates in solution in the same way as acetaldehyde, and show a similar acid catalysis in their dehydration. With furfural, on the other hand, no evidence of hydration was obtained.

PREPARATION OF MATERIALS

Commercial pure acetone was fractionally distilled once, and this product gave kinetic results identical with those obtained with acetone purified through the sodium iodide compound (table 1). Acetaldehyde was redistilled in an atmosphere of nitrogen directly before use. Some samples of acetaldehyde were found to contain traces of halogen acid, which was removed by a preliminary distillation from three times its volume of water. The acetaldehyde-water mixture was stored under nitrogen in a sealed flask to minimize oxidation, and 0·5 % of hydroquinone was usually added for the same purpose: this was shown to have no effect on the kinetic results.

The catalysts used were of 'Analar' grade or were subjected to recrystallization of fractional distillation. Such purification was repeated until a further stage caused no change in the relevant catalytic coefficient, and if possible different sources and different methods of purification were employed. The methods of purification and the melting-point or boiling-point of the product are given in table 3.

INTERPRETATION OF KINETIC RESULTS

Even with the strictest precautions it was found impossible to reduce the acetic acid concentration in the final mixture below about 1×10^{-4}M (corresponding to the oxidation of 0·05 % of the acetaldehyde present) and a more usual figure was 5×10^{-4}M.

The dehydration reaction took place at an appreciable rate in these solutions, and experiments in which acetic acid was added to the acetone showed that this catalysis was mainly due to their acetic acid content. The data are given in table 1 and plotted in figure 1. In some experiments the total acetic acid concentration was measured by titration, while in others the acetone solution was made up by weight and the acetaldehyde solution titrated. It will be seen that the reaction velocity is closely proportional to the total concentration, the scatter of the points being attributable to the titration errors in dilute solution. It will be shown later that the small intercept at $c = 0$ in figure 1 can be reasonably attributed to catalysis by water molecules.†

TABLE 1. CATALYSIS BY ACETIC ACID

c = total acetic acid concentration in moles/l. k = first-order velocity constant, minutes, decadic logarithms. Experiments marked with an asterisk were carried out with acetone purified through the sodium iodide compound.

10^3c	0·10	0·25	0·70	0·85	1·01	1·15	1·25	1·48	1·95	1·95
$100k$	0·674	0·748*	1·59*	1·55	2·01	2·50*	2·10	3·08	3·71	3·90*
10^3c	2·20	2·89	2·95	3·39	3·45	3·75	3·83	4·50	5·00	5·50
$100k$	4·35	5·85	5·65*	7·10	6·40	7·08*	7·93	8·55	9·00*	11·0

Although acetic acid could not be completely eliminated from the acetaldehyde solutions it was found that the acetic acid content of a given solution kept under nitrogen remained sensibly constant over 24 hr. even when 10 to 12 withdrawals

† After the work had been completed, data were published by Lauder (1948) on the slow changes of density and refractive index occurring when acetaldehyde and water are mixed in various proportions. He finds these changes to be kinetically of the first order, but the rates are highly irreproducible and appear to be greater at 0° than at 13·5° C. These anomalies are probably due to the presence of varying small amounts of acetic acid. The occurrence of a first-order change suggests that the equilibrium proportion of hydrate is a small one (as Lauder himself concludes from the observed refractive indices) and his hypothesis of two consecutive changes is unnecessary.

were made during this period. In making measurements with added catalysts, it was found that the plots of k against catalyst concentration were linear, different aldehyde solutions giving parallel lines with different intercepts on the k-axis. In order to minimize any error due to a progressive increase in acetic acid concentration, experiments were always done alternately with more concentrated and less concentrated catalyst solutions. Typical plots obtained in this way are shown in figure 2, and the catalytic constants given in this paper are all derived from plots of this kind.

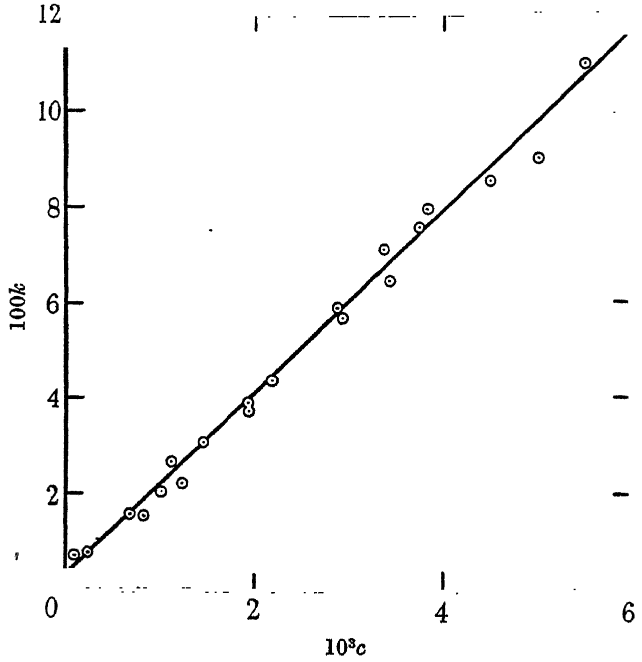


FIGURE 1

They clearly cannot be very accurate for catalysts like 1-nitropropane (cf. figure 2), where the change in velocity amounts only to about 15% in the most concentrated solutions, but for most catalysts they should not be in error by more than a few per cent.

In the experiments with saccharin as catalyst, where the maximum concentration which could be employed is less than 10^{-4} M, the observed velocities varied erratically, and the k-c plot makes an intercept on the c-axis at about $1{\cdot}5\times10^{-5}$ M (figure 3). This behaviour was traced to the adsorption of an appreciable amount of catalyst on the walls of the dilatometer, a phenomenon which has previously been observed in another catalytic reaction (Bell & Rybicka 1947). If the dilatometer was packed with thin glass rods all the velocities were decreased, and the intercept on the c-axis rose to about 4×10^{-5}. The adsorption process is not an instantaneous one, since the rate falls during the first few minutes of the reaction, this fall being absent if the dilatometer is first rinsed out with the solution to be used. The intercept of $1{\cdot}5\times10^{-5}$ M corresponds to about 3×10^{15} molecules per sq.cm. of the glass surface, which is a reasonable figure for a single layer. It seems, therefore, that the amount of catalyst adsorbed is roughly constant throughout, and that the catalytic constant can be taken as equal to the slope of the line drawn in figure 3. Adsorption is probably also detectable in the data for dichloracetic acid, but it will have no appreciable effect with any of the weaker acids, since the concentrations employed are much greater. On the other hand, with acids much stronger than dichloracetic the concentrations which can be used are so low that irreproducible results would be anticipated: this was in fact found to be the case for trichloracetic and picric acids.

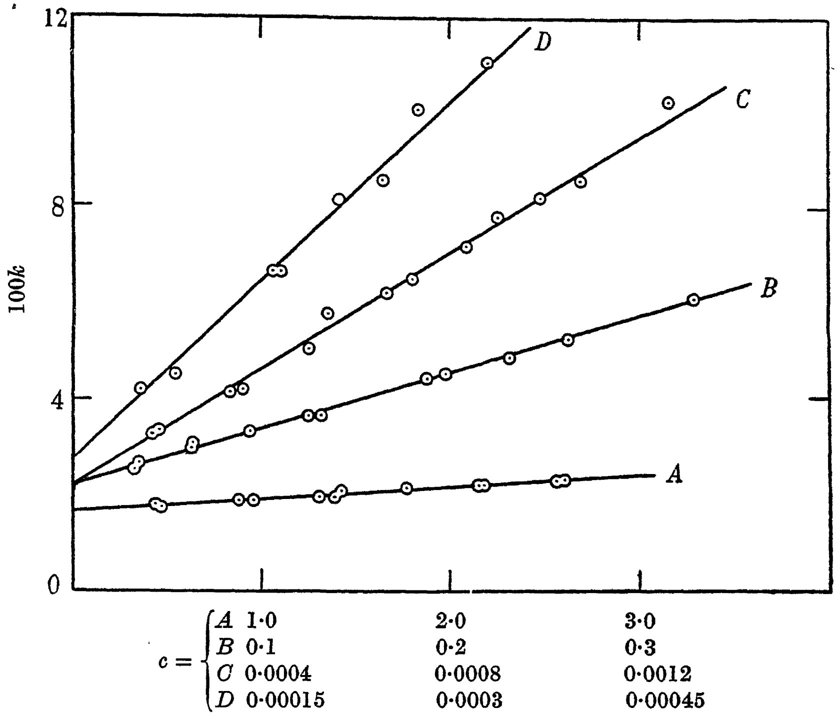


FIGURE 2. A = 1-nitropropane, B = acetoxime, C = m-toluic, D = cyanacetic.

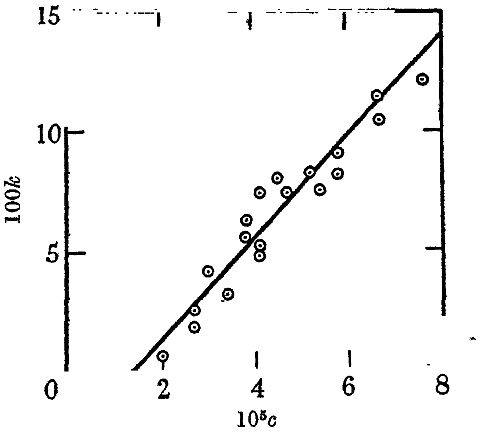


FIGURE 3. Saccharin.

The acids employed are slightly ionized in the reaction medium 92·5 % acetone + 7·5 % water, and approximate conductivity measurements show that acetic acid has a dissociation constant of about 10^{-8}, i.e. 1000 times smaller than in water. It is possible, therefore, that a proportion of the catalysis observed is due to hydrogen ions, though the linearity of the *k*-*c* plots shows that the proportion cannot be a large one. The point was investigated by making measurements with buffered solutions and it was found that the presence of about 10 % of the sodium salt of the acid was sufficient to repress the hydrogen ion effect completely. For example, $1{\cdot}9 \times 10^{-3}$ M acetic acid with the addition of 0, 3·3, 6·8 and 20 % of sodium acetate gave velocity constants of 0·0354, 0·0340, 0·0326 and 0·0328 respectively. Series of measurements were carried out with acetic, monochloracetic and dichloracetic acids, for which it was found that both buffered and unbuffered solutions give a linear relation between reaction velocity and acid concentration, the line for buffered solutions being somewhat lower. The results are summarized in table 2. Independent experiments showed that there was no detectable kinetic salt effect at the ionic strengths ($< 10^{-4}$) in these experiments.

TABLE 2. EFFECT OF BUFFERING

	acetic	monochloracetic	dichloracetic
number of experiments	14	9	12
maximum acid concentration	$3{\cdot}8 \times 10^{-3}$	$6{\cdot}3 \times 10^{-4}$	$1{\cdot}3 \times 10^{-4}$
buffer ratio (acid : base)	10·0	1·8	0·92
percentage reduction in rate: obs.	8 ± 5	13 ± 5	20 ± 5
calc.	4	11	24

The true catalytic constants for the undissociated acid molecules are of course those measured in the buffered solutions, and it would be desirable in principle to investigate buffered solutions for all the catalysts studied. There are, however, difficulties for acids in the pK range 3 to 7 because of the unavoidable presence of acetic acid: the equilibrium $CH_3COOH + X^- \rightleftharpoons CH_3COO^- + HX$ will assume an intermediate position, and the contribution of acetic acid catalysis to the observed rate can no longer be assumed constant. For acids with pK > 7 this difficulty disappears, since the acetic acid will be converted completely to acetate ion, but with such weak acids the concentrations are so high that there are appreciable kinetic salt effects in buffer solutions; moreover, the anions of very weak acids may be expected to act as basic catalysts, and such basic catalysis was in fact observed in solutions of sodium *p*-nitrophenate. Direct measurements with buffer solutions are therefore confined to acetic acid itself and the small number of acids with pK < 3. Fortunately the correction for hydrogen-ion catalysis can be calculated with sufficient accuracy, and is negligible in many cases.

Let k be the measured velocity in an unbuffered solution of concentration c of which k' is due to hydrogen ions. Then if dissociation constants in the reaction medium are proportional to those in water, we can write $k' = A\sqrt{(K_A c)}$, where K_A is the dissociation constant of the acid in water, and A a constant for all the catalyst solutions. It will be seen later that for the majority of the catalysts the observed catalytic constants in unbuffered solutions are related to the values of K_A by a relation of the form $k/c = GK_A^{0\cdot 55}$, G being a constant for all the catalysts. Combining the last two equations we find

$$k'/k = AG^{-\frac{1}{2}}k^{-\frac{1}{2}}K_A^{0\cdot 225}. \tag{1}$$

Equation (1) shows that the proportion of catalysis due to hydrogen ions increases with increasing acid strength, as found in table 2. It should also depend on the observed velocity k, but the data obtained are not sufficiently accurate to detect this with certainty over the small range of velocities ($k = 0\cdot 02$ to $0\cdot 1$) covered. This range was approximately the same for all the catalysts, and for the purpose of obtaining the true catalytic constant k_c from the observed slope of the k-c plot we can write equation (1) in the form $k_c'/k_c = BK_A^{0\cdot 225}$, where k_c' is the correction which must be subtracted from k_c to allow for the effect of hydrogen ions, and B is a constant for all the catalysts. The best value of B to fit the data in table 2 is $B = 0\cdot 4$, giving

$$k_c'/k_c = 0\cdot 4K_A^{0\cdot 225}. \tag{2}$$

The values calculated from (2) are given in table 2, and this relation was used throughout in correcting the observed values of k_c. The maximum correction is 24 % for dichloracetic acid, and most of the corrections made were less than 10 %. For acids with $pK > 6$ equation (2) predicts a correction < 2 %, and for such acids no correction was applied. None of the general conclusions of this paper would be modified by modifying the correction for hydrogen-ion catalysis in any reasonable manner, or by omitting it altogether.

DISCUSSION

The above data fully establish that the dehydration of acetaldehyde hydrate exhibits general catalysis by acids (of zero and positive charge) and bases (or zero and negative charge). The probable mechanism has already been given (p. 143), and the second step is likely to be the rate-determining one both for acid and for basic catalysis, since the first step in each case involves only the formation or fission of an O-H bond, which experience shows to be very rapid.

Table 3 gives data for the catalytic coefficients (k_c) of 63 uncharged acids, and it is of interest to compare these with their acid strengths as measured by their dissociation constants in water. In making such a comparison it is necessary to allow for statistical differences between the catalysts: thus if the acid has p equivalent detachable protons, and the corresponding base q equivalent points of attachment for a proton, we must compare k_c/p with qK_A/p (Brönsted 1928). We have adopted the following values:

$p = 1, q = 1$ monohydric phenols, oximes, violuric acid, phthalimide, saccharin, 2-nitropropane, nitrourea, nitrourethane.

$p = 2, q = 1$ dihydric phenols, chloral hydrate, nitroethane, 1-nitropropane.

$p = 1, q = 2$ carboxylic acids, benzoylacetone, dimedone.

$p = 3, q = 1$ nitromethane.

Some of these values are ambiguous, but none of the general conclusions would be affected by alternative choice. The values of K_A in table 3 are taken from Landolt-Bornstein *Tabellen* (1937), except those for benzoylacetone (Eidinoff 1945), dimedone (Schwarzenbach & Felder 1944) the nitro-paraffins (Turnbull & Maron 1943), and the aliphatic oximes (King & Marion 1944).

Figure 4 shows a plot of $\log_{10}(k_c/p)$ against $\log_{10}(qK_A/p)$ for all the carboxylic acids and phenols in table 3, the numbers in the figure being the ordinal numbers in the table. The straight line drawn in the figure represents the relation

$$\log_{10}\frac{k_c}{p} = 0{\cdot}54\log_{10}\frac{qK_A}{p} + 3{\cdot}75 \tag{3}$$

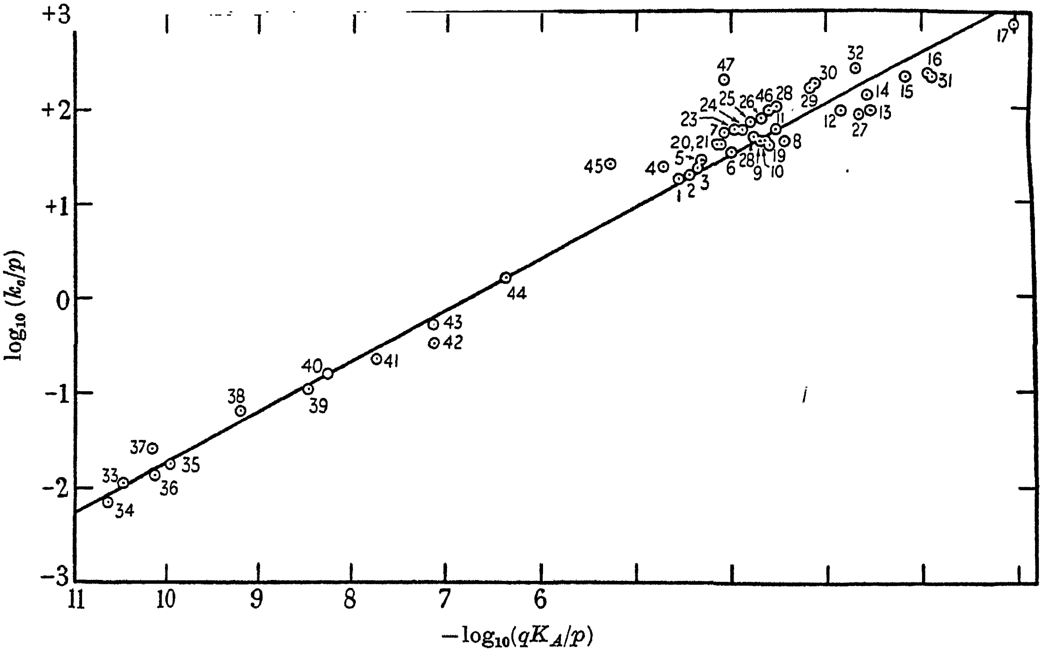


FIGURE 4

and the observed values of k_c agree with this relation within 0·3 logarithmic unit, with the exception of pentachlorophenol and 2.4-dinitrophenol, where the divergence is +0·6 and +0·8 respectively. (If the experimental values of k_c uncorrected for hydrogen-ion catalysis are used, an equally good straight line is obtained, of slope 0·55.) Closer examination reveals some further regularities: for example, the points referring to aliphatic acids and ortho-substituted benzoic acids lie somewhat below the line represented by equation (3), and those for meta- and para-substituted benzoic acids somewhat above it.

It would of course be desirable to use dissociation constants measured in the reaction medium (92·5 % acetone + 10 % water) in place of water. Since these are not available, it is of interest to compare the kinetic data with relative acid strengths in *n*-butanol, which has a similar dielectric constant. Data available for 22 of the acids used here (Wooten & Hammett 1935), and figure 5 shows a plot of $\log_{10}(k_c/p)$

TABLE 3. COLLECTED CATALYTIC CONSTANTS

k_c (corr.) = catalytic constant (decadic logarithms, mins.$^{-1}$, litres, moles^{-1}), corrected for hydrogen ion catalysis. c = catalyst concentration, moles/l. K_A = dissociation constant in water at 25° C.

acid	purification	number of experiments	maximum c	k_c (corr.)	K_A
	Aliphatic acids, and aromatic side-chain acids				
1. propionic	dist. b.p. 139·5 to 140·5° C.	7	5×10^{-3}	18·0	$1{\cdot}35\times10^{-5}$
2. acetic	A.R.	28	6×10^{-3}	19·2	$1{\cdot}76\times10^{-5}$
3. β-phenylpropionic	recryst. H_2O m.p. 48·5° C	5	3×10^{-3}	23·0	$2{\cdot}19\times10^{-5}$
4. trimethylacetic	dist. m.p. 34·0° C	17	3×10^{-3}	23·6	$9{\cdot}4\times10^{-6}$
5. crotonic	recryst. H_2O twice m.p. 71·8° C	7	3×10^{-3}	28·7	$2{\cdot}3\times10^{-5}$
6. phenylacetic	recryst. H_2O m.p. 76·4° C	14	2×10^{-3}	33·0	$4{\cdot}88\times10^{-5}$
7. cinnamic	recryst. aq. EtOH m.p. 133·1° C	11	2×10^{-3}	40·5	$3{\cdot}8\times10^{-5}$
8. formic	A.R.	13	2×10^{-3}	43·5	$1{\cdot}77\times10^{-4}$
9. β-chloropropionic	recryst. ligroin m.p. 39·5° C	13	1×10^{-3}	43·9	$1{\cdot}04\times10^{-4}$
10. diphenylacetic	recryst. H_2O m.p. 145·8° C	7	2×10^{-3}	44·2	$1{\cdot}15\times10^{-4}$
11. glycollic	—	12	1×10^{-3}	56·6	$1{\cdot}54\times10^{-4}$
12. phenoxyacetic	recryst. H_2O m.p. 98·1° C	8	9×10^{-4}	92	$7{\cdot}6\times10^{-4}$
13. bromacetic	recryst. ligroin m.p. 49·1° C	10	6×10^{-4}	129	$1{\cdot}38\times10^{-3}$
14. chloracetic	A.R.	24	6×10^{-4}	146	$1{\cdot}51\times10^{-3}$
15. cyanacetic	recryst. $CHCl_3$ m.p. 66·8° C	9	3×10^{-4}	218	$3{\cdot}56\times10^{-3}$
16. phenylpropiolic	recryst. H_2O m.p. 135·1° C	12	4×10^{-4}	225	$5{\cdot}9\times10^{-3}$
17. dichloracetic	dist. b.p. 193° C	17	2×10^{-4}	773	$5{\cdot}0\times10^{-2}$
	Aromatic acids				
18. o-methoxybenzoic	recryst. H_2O m.p. 101·0° C	12	2×10^{-3}	27·6	$8{\cdot}2\times10^{-5}$
19. o-toluic	recryst. H_2O m.p. 103·1° C	12	2×10^{-3}	38·3	$1{\cdot}22\times10^{-4}$
20. p-hydroxybenzoic	recryst. EtOH + xylene m.p. 211·1° C	12	2×10^{-3}	38·8	$2{\cdot}85\times10^{-5}$

TABLE 3 *continued*

acid	purification	number of experiments	maximum c	k_c (corr.)	K_A
Miscellaneous acids					
21. *p*-methoxybenzoic	recryst. H_2O m.p. 181·7° C	13	2×10^{-3}	40·1	$3{\cdot}38\times10^{-5}$
22. *m*-hydroxybenzoic	recryst. aq. EtOH m.p. 198·6° C	13	1×10^{-3}	47·6	$8{\cdot}33\times10^{-5}$
23. *p*-toluic	recryst. H_2O m.p. 177·6° C	13	2×10^{-3}	50·4	$4{\cdot}33\times10^{-5}$
24. *m*-toluic	recryst. H_2O m.p. 111·0° C	14	1×10^{-3}	57·2	$5{\cdot}32\times10^{-5}$
25. benzoic	recryst. toluene m.p. 122·2° C	13	1×10^{-3}	59·2	$6{\cdot}6\times10^{-5}$
*26. *p*-chlorobenzoic	recryst. aq. EtOH m.p. 236·5° C	13	1×10^{-3}	77·1	$1{\cdot}04\times10^{-4}$
*27. *o*-chlorobenzoic	recryst. aq. EtOH m.p. 141·0° C	9	8×10^{-4}	85·1	$1{\cdot}20\times10^{-3}$
*28. *m*-chlorobenzoic	recryst. aq. EtOH m.p. 153·2° C	21	7×10^{-4}	101	$1{\cdot}5\times10^{-3}$
29. *m*-nitrobenzoic	recryst. aq. EtOH m.p. 140·0° C	21	4×10^{-4}	152	$3{\cdot}48\times10^{-4}$
30. *p*-nitrobenzoic	recryst. aq. EtOH m.p. 235·5° C	13	5×10^{-4}	159	$3{\cdot}93\times10^{-4}$
31. *o*-nitrobenzoic	recryst. H_2O m.p. 145·5° C	16	2×10^{-4}	221	$6{\cdot}5\times10^{-3}$
32. salicylic	A.R.	20	3×10^{-4}	250	$1{\cdot}06\times10^{-3}$
Phenols					
33. thymol	dist. m.p. 51·0° C	16	0·7	0·0112	$3{\cdot}2\times10^{-11}$
†34. hydroquinone	recryst. H_2O m.p. 170° C	15	0·2	0·013	$4{\cdot}5\times10^{-11}$
35. phenol	dist. m.p. 41·0° C	24	2·2	0·0181	$1{\cdot}06\times10^{-10}$
36. resorcinol	recryst. benzene twice m.p. 109·5° C	14	0·9	0·026	$1{\cdot}55\times10^{-10}$
37. pyrocatechol	recryst. benzene twice m.p. 104·3° C	15	1·0	0·049	$1{\cdot}4\times10^{-10}$
38. *p*-chlorophenol	dist. twice b.p. 216–217° C	11	0·8	0·061	$6{\cdot}6\times10^{-10}$
39. *o*-chlorophenol	dist. twice b.p. 174–175° C	14	0·7	0·112	$3{\cdot}2\times10^{-9}$
‡40. *m*-nitrophenol	recryst. H_2O, ligroin m.p. 95·7° C	35	0·4	0·16	$5{\cdot}3\times10^{-9}$
§41. 2.4-dichlorophenol	dist. 4 times b.p. 209–210° C	6	0·3	0·225	$1{\cdot}8\times10^{-8}$
42. *o*-nitrophenol	recryst. EtOH m.p. 45·7° C	17	0·2	0·334	$6{\cdot}8\times10^{-8}$
43. *p*-nitrophenol	recryst. H_2O m.p. 113·3° C	14	0·2	0·52	$6{\cdot}75\times10^{-8}$
44. 2.4.6-trichlorophenol	recryst. benzene twice m.p. 67·7° C	16	6×10^{-2}	1·53	$3{\cdot}9\times10^{-7}$
45. pentachlorophenol	recryst. benzene 6 times m.p. 189° C	13	6×10^{-3}	24·0	$5{\cdot}5\times10^{-6}$
46. 2.6-dinitrophenol	recryst. H_2O m.p. 62·3° C	13	1×10^{-3}	91	$2{\cdot}6\times10^{-4}$
47. 2.4-dinitrophenol	recryst. EtOH 4 times m.p. 112° C	22	6×10^{-4}	183	$8{\cdot}1\times10^{-5}$

48.	benzophenone oxime	recryst. EtOH twice m.p. 142° C	12	0·3	* 0·053	5 $\times 10^{-12}$
49.	acetophenone oxime	recryst. H_2O twice m.p. 58·2° C	14	0·4	0·066	3·3 $\times 10^{-12}$
50.	diethyl ketoxime	dist. *in vacuo*; stood 3 days over solid Na_2CO_3; redist. *in vacuo* b.p. 80·5° C (21 mm.)	13	0·2	0·104	2·5 $\times 10^{-13}$
51.	acetoxime	recryst. once ligroin, twice acetone, m.p. 60·8° C	15	0·4	0·118	3·8 $\times 10^{-13}$
‖52.	chloral hydrate	recryst. CS_2 twice m.p. 51·0° C	16	0·4	0·033	1 $\times 10^{-11}$
¶53.	benzoylacetone (enol)	recryst. Et_2O 3 times m.p. 56·5° C	11	0·3	0·0088	5·9 $\times 10^{-9}$
¶54.	dimedone	recryst. H_2O m.p. 147·5° C	8	4×10^{-2}	1·40	5·9 $\times 10^{-6}$
55.	phthalimide	recryst. Et_2O twice m.p. 232·8° C	14	4×10^{-2}	0·126	5 $\times 10^{-9}$
56.	violuric	recryst. H_2O dec. 210° C	10	4×10^{-2}	17·3	2·7 $\times 10^{-5}$
57.	nitrourea	recryst. Et_2O 3 times subl. 142° C (dec.)	18	4×10^{-3}	23·2	7·0 $\times 10^{-5}$
58.	nitrourethane	ether soln. pptd. 3 times ligroin m.p. 64·0° C	12	2×10^{-3}	36·1	5·3 $\times 10^{-4}$
59.	saccharin	recryst. EtOH twice m.p. 226° C	19	8×10^{-5}	2220	2·5 $\times 10^{-2}$
60.	nitromethane	dist. *in vacuo*; stood 6 days over solid Na_2CO_3; dist. *in vacuo* b.p. 42·0–42·5° C (80 mm.)	15	4·3	0·00084	5·8 $\times 10^{-10}$
61.	1-nitropropane	as above b.p. 36·5–37·5° C (18 mm.)	14	2·6	0·0026	1·05 $\times 10^{-9}$
62.	nitroethane	as above b.p. 29·5–30·5° C (24 mm.)	12	3·3	0·0028	2·5 $\times 10^{-9}$
63.	2-nitropropane	as above b.p. 28–29° C (18 mm.)	14	2·6	0·0038	1·8 $\times 10^{-8}$

[*] For these catalysts the k-c plot showed some falling off from linearity. k_c was obtained from the initial gradient.

[†] Measurements were limited by the small solubility of these catalysts and the resulting values of k_c are not very accurate.

[‡] At concentrations greater than $c = 0·1$ the k-c plot became curved. The addition of 0·4M menthol; naphthalene or *m*-dinitrobenzene (which do not catalyze the reaction) increased the rate in the curved region of the plot, but did not affect the linear part, which was used to derive k_c.

[§] The result for this catalyst may be inaccurate, since it is difficult to separate 2.4-dichlorophenol from the 2.5 and 2.6 isomers, and the fourth distillation produced an increase of 30 % in k_c.

[‖] Experiments showed that no volume change accompanies the dissolution of chloral hydrate in acetone, and acetic acid is the only solvent in which it is known to dissociate appreciably.

[¶] The percentage enolization of benzoylacetone in the medium used here was measured by Mr E. Gelles in this laboratory, using a bromination method (cf. Eidinoff 1945), giving 86·3 %. The value given for k_c is obtained by dividing the observed value by 0·86, which neglects any small catalytic effect of the keto-form. It was not found possible to estimate the enol content of solutions of dimedone (5.5-dimethyl cyclohexane-1.3-dione) by this method, since the bromoketone first formed undergoes rapid further bromination. However, aqueous solutions of dimedone contain 95 % enol (Schwarzenbach & Felder 1944), and we have assumed our solutions to contain 100 % enol, since the change from water to 92·5 % acetone raises the enol content of benzoylacetone from 34 to 86 %. The k-c plots for these two catalysts showed a slight curvature, which may indicate some reaction between catalyst and acetaldehyde: however, the volume change and the course of single reactions were quite normal, and no change was visible in the solutions after 2 hr., although dimedone is used as a reagent for aldehydes.

against $\log_{10}(qK'_A/p)$, where K'_A is the acid strength in *n*-butanol relative to benzoic acid.

The general picture is very similar to that in figure 4, the broken line representing the mean line for all the catalysts, and the full lines the separate relations for aliphatic and ortho-substituted benzoic acids (lower line) and for meta- and para-substituted acids (upper line). The point for 2:4-dinitrophenol now lies close to the mean line, and it seems likely that its discrepant position in figure 4 is due to an abnormal change in acid strength in passing from water to 92·5 % acetone: the same may be true for pentachlorophenol. Apart from this the data for *n*-butanol (which represent a particular case of the parallelism commonly found between relative acid strengths in different solvents, cf. Bell 1941, p. 108) justify the use of the more accessible dissociation constants in water.

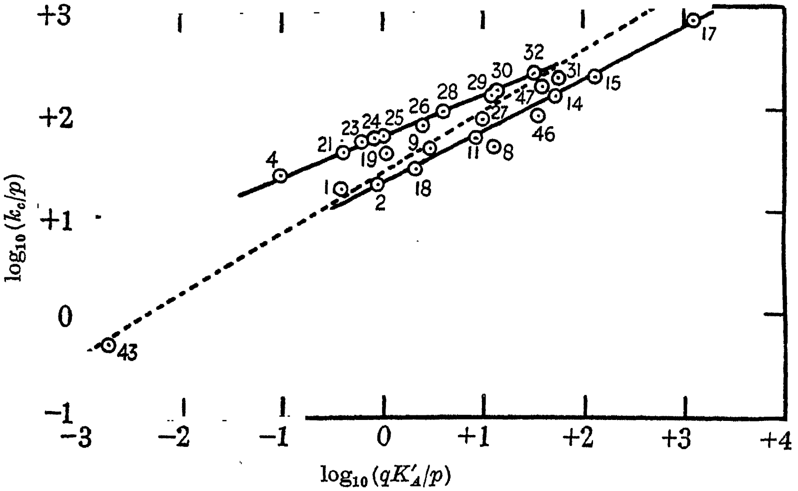


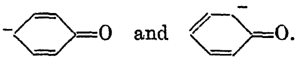
FIGURE 5

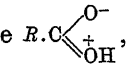
We shall return later to the detailed features of figure 4, but shall consider first the position of the sixteen miscellaneous acids at the end of table 3; some of which show considerable deviations from equation (3): these are listed in table 4.

TABLE 4. MISCELLANEOUS ACIDS

acid	logarithmic deviation from equation (3)	acid	logarithmic deviation from equation (3)
benzophenone oxime	+ 1·2	phthalimide	− 0·05
acetophenone oxime	+ 1·3_5	violuric acid	+ 0·05
diethyl ketoxime	+ 2·1	nitrourea	− 0·1
acetoxime	+ 2·1	nitrourethane	− 0·3_5
chloral hydrate	+ 0·7	nitromethane	− 1·4
benzoylacetone	− 1·4	1-nitropropane	− 1·5
dimedone	− 1·1	nitroethane	− 1·7
saccharin	+ 0·5_5	2-nitropropane	− 1·8_5

Before seeking a structural reason for these deviations we must consider why the carboxylic acids and the phenols conform to the same relation, in spite of structural differences. In the anion of a carboxylic acid the change is divided equally between two oxygen atoms, so that ionization involves some spread of charge away from the site of the proton. There is no corresponding change in the ionization of phenols as conventionally written, but this no doubt exaggerates the difference between the two classes. In the first place, the acidity of the phenols (some 10 powers of 10 greater than that of the alcohols) implies that the negative charge in the anion must be partly distributed over the benzene ring, which may be represented by participating structures such as



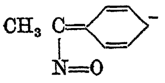
In the second place the undissociated carboxyl group has a considerable resonance energy, involving the structure  $R.C\begin{smallmatrix}O^-\\ \overset{+}{O}H\end{smallmatrix}$, so that the change in structure is not so great as might be supposed. The net result is that the spread of charge on ionization is not very different in degree for carboxylic acids and for phenols.

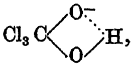
The first seven acids in table 4 all contain a proton attached to an oxygen atom. and can be legitimately compared with the phenols and carboxylic acids. In the four ketoximes there is no formal possibility of structural change on ionization and the negative charge must remain concentrated chiefly on the one oxygen atom. They are correspondingly 10 to 100 times more active than phenols of the same dissociation constant, and the same effect appears to a less marked extent with chloral hydrate. In the two β-diketones, on the other hand, the charge on the anion is shared between two oxygen atoms, e.g. for benzoylacetone,

$$\{C_6H_5.C(O^-):CH.CO.CH_3,\ C_6H_5.CO.CH:C(O^-).CH_3\}.$$

The structural change in the ionization of the enol thus resembles that in the carboxylic acids, but now involves a wider spread of charge: correspondingly, the catalytic constants are about 20 times smaller than equation (3) predicts. The next five acids in table 4, containing an acidic $>$NH group, do not show any large deviations, and since both the molecules and ions can be given a large number of bond assignments, no general conclusions can be drawn. Finally, the four nitroparaffins which conclude table 4 are 25 to 70 times less effective than anticipated. Although it is not entirely legitimate to compare them with —OH acids, it is reasonable to relate this low activity to the fact that, when a nitroparaffin ionizes, the negative charge is shifted almost completely away from the carbon atom and resides in the NO_2 group.

The large effects considered above provide good evidence for a general relation between the catalytic power of an acid of given strength and the extent to which a charge-shift takes place when it ionizes. We shall now examine how far some of the smaller differences can be accounted for along the same lines, though such considerations are bound to be more speculative. For example, the two aromatic ketoximes (table 4) are not such active catalysts as the aliphatic ones although they have ten times the acid strength: this can be related to the possibility of structures such as

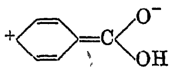


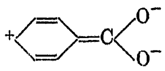
in the anion, which serve to shift some of the negative charge away from the oxygen atom. Such structures will not be as important as the corresponding ones in phenol so that the aromatic ketoximes still show a positive deviation from equation (3). The still smaller positive deviation shown by chloral hydrate may be due to hydrogen bonding in the anion,  which will have the effect of shifting some of the negative charge on to the other oxygen atom.

In order to explain the distinction found (cf. figure 5) between benzoic acids and its meta- and para-derivatives on the one hand and aliphatic acids (including aromatic acids with the carboxyl group in the side-chain) on the other, we must consider the factors which determinate the strength of aromatic acids. Considering the series

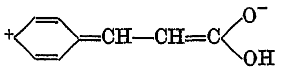
$$C_5H_5COOH\ (K_a = 6{\cdot}6\times 10^{-5}),\ C_6H_5CH_2COOH\ (K_a = 4{\cdot}9\times 10^{-5}),$$
$$C_6H_5CH_2CH_2COOH\ (K_a = 2{\cdot}2\times 10^{-5})$$

there is clearly some factor present reducing the strength of benzoic acid, which on a normal inductive basis would be expected to have K_a at least 2×10^{-4} (cf. formic acid, $K_a = 1{\cdot}8\times 10^{-4}$). This factor is probably the contribution of structures like (*a*) (and similar structures with a positive change in the ortho-position) to the mesomeric state of the molecule.

(*a*) 

(*b*) 

The corresponding structure (*b*) for the anion is relatively improbable, since it involves the loss of the normal resonance energy of a carboxylate ion, and the proximity of two large negative charges: hence the effect of such structures is to decrease the strength of the acid by stabilizing the molecule relative to the ion. This is the reverse effect to the one we have already considered (the stabilization of the anion by types of mesomerism which cannot occur in the acid molecule), and we should expect it to have the opposite kinetic effect, i.e. to increase the reaction velocity. This accounts for the fact that the aromatic acids are better catalysts than aliphatic acids of the same strength, and it also explains the inclusion of cinnamic acid among the aromatic acids, since structures of the type



are possible. The same picture accounts for the high dissociation constants of the ortho-substituted benzoic acids and the fact that they fall together with the aliphatic rather than the aromatic acids in figures 4 and 5. A structure such as (*a*) above demands that the COOH-group shall lie in the plane of the ring, and this will be sterically impossible if there is an ortho-substituent: hence these structures are unable to stabilize the undissociated molecule, and the substance behaves kinetically as an aliphatic rather than as an aromatic acid. Accordingly it is to be expected that the position of the ortho-substituted acids on the Brönsted plot would be below the meta- and para-aromatic acid line, but not below the aliphatic line, which represents a lower limit. Reference to figure 5 shows this to be the case. The only exception to this rule is salicylic acid, where the position is complicated by hydrogen-bonding with the OH-group.

The experimental data provide a rough estimate of the catalytic power of the species H_2O. The small intercept in figure 2 corresponds to a velocity of about 4×10^{-3} in a 4M solution of water, giving $k_c(H_2O) \sim 10^{-3}$. (It is clear from the experiments with amine bases that any basic catalysis by the water molecule will be negligible.) The usual value given for the acid strength of the H_2O molecule is $K_w/[H_2O] = 10^{-14}/55{\cdot}5 = 2 \times 10^{-16}$. Putting this value in equation (3) with $p = 2$, $q = 1$, we obtain $k_c = 2{\cdot}5 \times 10^{-5}$, which is 40 times smaller than the observed value, and would not be detectable experimentally. The value 2×10^{-16} probably underestimates the acid strength of the H_2O molecule, and it has been estimated (Bell 1943) that the true value lies somewhere between 2×10^{-16} and 8×10^{-14}. However, even using the higher of these two figures the calculated value is 5×10^{-4}, and it seems certain that the catalytic power of the H_2O molecule is considerably greater than equation (3) predicts. This is just what would be expected from the above considerations, since there are clearly no possibilities of mesomeric charge shift in the ion OH^-.

Our experimental results demonstrate that for uncharged acids of equal strength the velocity of protolytic change is decreased by the occurrence of a mesomeric change when the anion is formed, and increased by the existence of mesomerism in the acid which cannot occur in the anion. In particular, in an acid catalyzed reaction a series of catalysts, for which the variations in strength are mainly determined by inductive effects, will conform to a single Brönsted relation, while the occurrence of variable mesomeric effects will lead to deviations from such a relation. The same regularities will of course hold for the velocity of reverse process of the transfer of a proton to the anion (e.g. basic anion catalysis), since the position of equilibrium depends only on the strength of the acid. We shall now show that these regularities can be accounted for in terms of molecular potential energy curves. In figure 6, let $ABCD$ be the dissociation curve for an acid HX with no mesomeric change in the anion. The dissociation curve of another acid YH, having the same strength as XH, but with an anion stabilized by charge shift, will be represented by $ABC'D$, lying above the first curve. If the charge-shift were inhibited completely the acid YH would be weaker, dissociating to the level D' where DD' is the stabilization energy of the anion Y^-. At an intermediate internuclear distance the charge shift in the actual acid will be incomplete, so that $C''C' < C''C < D'D$. The converse argument will apply if mesomerism is present in the acid instead of the anion. This argument can be extended to the velocities of transfer of a proton from the acids XH and YH to a base S (e.g. a substrate in a catalyzed reaction). In figure 7, curve $ABCDE$ represents the energy changes during the reaction between XH and S, B being the equilibrium position for $XH + S$, D that for $X^- + HS^+$, and C the transition state. The heat of reaction Q and the activation energy E are shown on the diagram. The corresponding curve for YH is $ABC'DE$, where the heat of reaction is unchanged (since the acids are of equal strength) but the activation energy is increased by an amount DE, which will be less than the stabilization energy of Y^-. The value of DE clearly depends on the nature of the transition state: if it is close to the initial state, *DE* will be small, while if it approximates to the final state, *DE* may be nearly equal to the stabilization energy. These diagrams show that the occurrence of structural changes on ionization is associated with variations in the shapes of the potential energy curves. This explains why such changes cause deviations from the Brönsted relation, since the theoretical treatment of this relation assumes a series of curves which differ in relative position, but have the same shape (Horiuti & Polanyi 1935; Bell 1936).

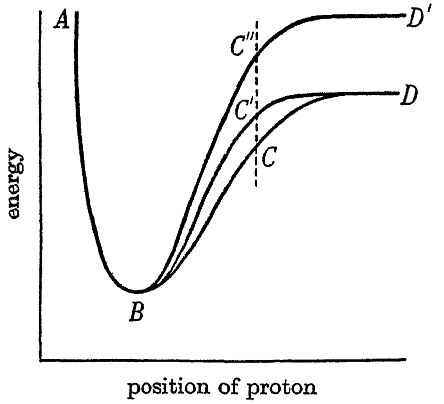


FIGURE 6

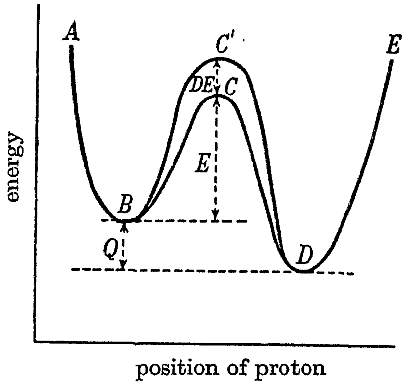
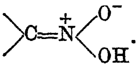


FIGURE 7

For the nitroparaffins it is possible to attribute at least a part of the observed reaction velocity to catalysis by the small equilibrium concentration of the aci-form  $\gt C=\overset{+}{N}\begin{matrix}\diagup O^- \\ \diagdown OH\end{matrix}$. If we use the concentrations and dissociation constants of the aci-form measured in aqueous solution (Turnbull & Maron 1943) the velocities predicted by equation (3) are in fact fairly close to the observed values. These values therefore represent upper limits for the catalytic effect of the nitro-forms, which may be much smaller, or even zero. There is no evidence that such tautomeric forms exist for any of the other catalysts used, and they can be assigned separate catalytic effects only if they react with the substrate through different transition states. If any corrections for tautomerism were necessary, their effect would always be to increase the negative deviations found, so that the structural arguments advanced would still be valid.

Our thanks are due to Imperial Chemical Industries (Billingham) for the gift of pure nitroparaffins, and to the Department of Scientific and Industrial Research for a maintenance grant to one of us (W. E. C. H.).

REFERENCES

Alquier, R. 1943 *Bull. Soc. Chim. Fr.* (5), **10**, 197.
Bell, R. P. 1934 *Proc. Roy. Soc.* A, **143**, 377.
Bell, R. P. 1936 *Proc. Roy. Soc.* A, **154**, 414.
Bell, R. P. 1940 *Nature*, **146**, 166.
Bell, R. P. 1941 *Acid-Base Catalysis*. Oxford University Press.

Bell, R. P. 1943 *Trans, Faraday Soc.* **39**, 253.
Bell, R. P. 1947 *Chem. Soc. Quart. Rev.* **1**.
Bell, R. P. & Baughan, E. C. 1937 *J. Chem. Soc.* p. 1947.
Bell, R. P. & Hirst, J. P. H. 1939 *J. Chem. Soc.* p. 1777.
Bell, R. P. & Rybicka, S. 1947 *J. Chem. Soc.* p. 24.
Booth, V. H. & Roughton, F. J. W. 1938 *Biochem. J.* **32**, 2049.
Brönsted, J. N. 1928 *Chem. Rev.* **5**, 322.
Brönsted, J. N. & Bell, R. P. 1931 *J. Amer. Chem. Soc.* **53**, 2478.
Brönsted, J. N. & Guggenheim, E. A. 1925 *J. Amer. Chem. Soc.* **49**, 2554.
Brönsted, J. N. & Pedersen, K. J. 1924 *Z. phys. Chem.* **108**, 185.
Brown, H. T. & Pickering, P. S. U. 1897 *J. Chem. Soc.* **71**, 774.
Cohn, M. & Urey, H. C. 1938 *J. Amer. Chem. Soc.* **60**, 679.
Colles, W. M. 1906 *J. Chem. Soc.* **89**, 1246.
Dieckmann, W. 1916 *Ber. dtsch. chem. Ges.* **49**, 2213.
Dieckmann, W. 1917 *Ber. dtsch. chem. Ges.* **50**, 1375.
Eidinoff, M. L. 1945 *J. Amer. Chem. Soc.* **67**, 1072.
Guggenheim, E. A. 1926 *Phil. Mag.* (7), **1**, 538.
Herbert, J. B. M. & Lauder, I. 1938 *Trans. Faraday Soc.* **34**, 432, 1219.
Herold, W. & Wolf, K. L. 1931 *Z. phys. Chem.* B, **12**, 165.
Homfray, I. F. 1905 *J. Chem. Soc.* **87**, 1435.
Horiuti, J. & Polanyi, M. 1935 *Acta Physicochim. U.R.S.S.* **2**, 505.
King, C. V. & Marion, A. P. 1944 *J. Amer. Chem. Soc.* **66**, 977.
Lauder, I. 1948 *Trans. Faraday Soc.* **44**, 729.
Lowry, T. M. & Smith, G. F. 1927 *J. Chem. Soc.* p. 2539.
McKenzie, A. & Mitchell, A. G. 1929 *Biochem. Z.* **208**, 456.
McKenzie, A. & Ritchie, P. D. 1931 *Biochem. Z.* **231**, 412; **237**, 1.
McKenzie, A. & Ritchie, P. D. 1932 *Biochem. Z.* **250**, 376.
Perkin, W. H. 1887 *J. Chem. Soc.* **51**, 808.
Ramsay, W. & Young, S. 1886 *Phil. Trans.* **177**, 71.
Schaefgen, J. R., Verhoek, F. H. & Newman, M. S. 1945 *J. Amer. Chem. Soc.* **67**, 253.
Schwarzenbach, G. & Felder, E. 1944 *Helv. chim. Acta,* **27**, 1701.
Schou, S. A. 1926 *C.R. Acad. Sci., Paris,* **182**, 965.
Tong, L. K. J. & Olson, A. R. 1941 *J. Amer. Chem. Soc.* **63**, 3406.
Turnbull, D. & Maron, S. H. 1943 *J. Amer. Chem. Soc.* **65**, 212.
Walker, E. 1944 'Formaldehyde', *Amer. Chem. Soc. Monogr.* Ch. III.
Wooten, L. A. & Hammett, L. P. 1935 *J. Amer. Chem. Soc.* **57**, 2289.

Some remarks on integral equations with kernels: $L(\xi_1 - x_1, \ldots, \xi_n - x_n; \alpha)$

By G. Kreisel, *Trinity College, University of Cambridge*

(*Communicated by E. C. Bullard, F.R.S.—Received* 6 *January* 1948—
Revised 8 *December* 1948)

The paper discusses integral equations of the type

$$k(\xi_1 \ldots \xi_n) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \psi(x_1 \ldots x_n)\, L(\xi_1 - x_1 \ldots \xi_n - x_n;\, \alpha)\, dx_1 \ldots dx_n,$$

where L is L-integrable, and ψ and k are bounded.

Since rapidly oscillating ψ have a small k, and since measurements of k are necessarily uncertain within non-zero limits of experimental error, very different ψ are consistent with any given set of measurements of k. Thus ψ is not determined by measurements of k.

Instead of ψ, partial information about ψ that is not sensitive to rapid oscillations of ψ, can be obtained from k. In the present paper we consider *smoothed versions* of ψ, and their applications to gravity survey and the theory of surface waves.

(1) For given L we construct normalized smoothing functions $\mu(x_1 \ldots x_n)$, so that

$$\bar{\psi}(x_1 \ldots x_n) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \psi(u_1 \ldots u_n)\, \mu(x_1 - u_1 \ldots x_n - u_n)\, du_1 \ldots du_n$$

can be calculated from measurements of $k(\xi_1 \ldots \xi_n)$. The method is applied to gravity survey, where the distribution ψ of masses on a plane Σ' is to be calculated from the normal force k on another (parallel) plane Σ.

(2) By studying suitable smoothing functions we get *a lower bound for the maximum modulus of the functions* ψ which are consistent with given experimental values of k. The bound is large if k is known to vary rapidly. The bound is also large if α is large and if the Fourier transform of L tends to 0 when $\alpha \to \infty$. The results are applied to gravity survey where now we consider the normal force on Σ due to masses of bounded density distributed in the space below Σ', where Σ' is itself below Σ. If the normal force is not uniform the distance between Σ and Σ' must not be too large, the estimate depending on the bound for the density.

Also fairly general conditions are imposed on ψ so that ψ is approximately determined by measurements of k, and an example from the theory of propagation in dispersive media is given where such conditions may be justified.

The gist of the paper is contained in theorems A and B.

I. INTRODUCTION

1. In the present paper we consider integral equations of the form[†]

$$k(\xi_1 \ldots \xi_n) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \psi(x_1 \ldots x_n)\, L(\xi_1 - x_1 \ldots \xi_n - x_n;\, \alpha)\, dx_1 \ldots dx_n, \tag{1·1}$$

where k is measured and ψ is to be determined.

[†] Such equations occur frequently. The following are some important examples:

(*a*) $n = 2$. Suppose we have a distribution of sources $\psi(x, y)$ on an infinite plane (e.g. gravitational masses or electric charges) which have an effect $k(\xi, \eta)$ on a parallel plane (the effect being, for example, the normal force or the potential); provided the fields of these sources may be linearly superposed, and the effect at (ξ, η) of unit source at (x, y) depends only on $(\xi - x, \eta - y)$ and not on (x, y), then the connexion between $k(\xi, \eta)$ and $\psi(x, y)$ will be of the form (1·1).

Throughout we assume that k and ψ are bounded, and that the kernel L is L_1, where, as usual, we say that a function f belongs to the class L_p if

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} | f(x_1 \ldots x_n) |^p \, dx_1 \ldots dx_n < \infty.$$

Occasionally other restrictions will be imposed on ψ.

2. Observe that a k which satisfies (1·1) is necessarily continuous since

$$\begin{aligned} &| k(\xi_1 + \epsilon_1 \ldots \xi_n + \epsilon_n) - k(\xi_1 \ldots \xi_n) | \\ &\qquad < \max | \psi | \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} | L(x_1 + \epsilon_1 \ldots x_n + \epsilon_n) - L(x_1 \ldots x_n) | \, dx_1 \ldots dx_n, \end{aligned}$$

which, by the theory of the Lebesgue integral, is small if the ϵ are small. Therefore, the values of k cannot be assumed to be discrete, and can only be measured to within certain non-zero limits of experimental error.[‡] Consider then an application of (1·1) in which any ψ of the class $\mathscr{F}$ must be expected to occur; in general, a whole range of ψ will be consistent with a given set of measurements of k, namely, any ψ of $\mathscr{F}$ whose transform

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \psi(x_1 \ldots x_n) \, L(\xi_1 - x_1 \ldots \xi_n - x_n) \, dx_1 \ldots dx_n \tag{2·1}$$

lies within the limits of uncertainty of k. Consequently the measurements can only be said to determine those results about ψ which are true of every ψ of this range. We therefore generalize the problem of para. 1, and propose to establish results of this kind.

Indeterminacy of ψ

3. The clue to finding results which are determined by measurements of k is given by the proof of the theorem: the values of ψ itself are not even approximately determined by measurements of k, however small their uncertainty may be.

To formulate the theorem precisely, assume that the results of measurements of k are presented in the following form:

If k is measured at $(\xi_{1i} \ldots \xi_{ni})$, $1 \leqslant i \leqslant N$, we put limits c_i on the indetermination of k in a neighbourhood I_i of $(\xi_{1i} \ldots \xi_{ni})$, so that

$$| k(\xi_1 \ldots \xi_n) - k_i | < c_i \quad (c_i > 0), \qquad \text{when} \quad (\xi_1 \ldots \xi_n) \text{ is in } I_i. \tag{3·1}$$

(*b*) $n = 1$. Suppose we have an instrument whose response $k(\tau)$ at time τ to an input $\psi(t)$ is given by

$$k(\tau) = \int_{-\infty}^{\tau} \psi(t) \, L(\tau - t) \, dt,$$

that is, the response at τ is calculated from the input before τ. Then we have an integral equation of type (1·1), where $L(u) = 0$ for $u < 0$.

(*c*) Suppose we want to calculate the disturbance at one point of a vibrating medium at one time from measurements of the disturbance at another point of the medium; in general, if the differential equations for the disturbance are linear and do not depend on the origin of time the integral equation connecting the disturbances at the two points will be of form (1·1).

[‡] Observe in passing that the kernel L, which characterizes the physical problem, is usually obtained theoretically. We therefore assume that it can be calculated arbitrarily closely, and has no experimental uncertainty.

Also in the space I_0 outside $\sum_1^N I_i$ we put a bound on k:

$$\overline{bd}\,|\,k(\xi_1 \dots \xi_n)\,| < c_0 \quad \text{for} \quad (\xi_1 \dots \xi_n) \text{ in } I_0. \tag{3·2}$$

We adopt the convention that the I_i, $1 \leqslant i \leqslant N$, are closed and do not overlap. For short, we shall say k, or ψ, is consistent with $\{k_i, c_i\}$ if k or the transform (2·1) of ψ satisfies (3·1) and (3·2).

Now we prove:

Suppose we have a bounded ψ_0 consistent with $\{k_i, c_i\}$. Then also

$$\psi_1 = \psi_0(x_1 \dots x_n) + A\sqrt{(2\pi)^{-n}}\, e^{i(p_1x_1+\dots+p_nx_n)}$$

is consistent with $\{k_i, c_i\}$, where A may be taken arbitrarily large provided only that $p_1^2+\dots+p_n^2$ is also sufficiently large. Observe that

$$\sqrt{(2\pi)^{-n}}\, A\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L(\xi_1-x_1 \dots \xi_n-x_n)\, e^{i(p_1x_1+\dots+p_nx_n)}\, dx_1 \dots dx_n = A\Lambda\, e^{i(p_1\xi_1+\dots+p_n\xi_n)},$$

where Λ is the Fourier transform of L; since L is L_1, by the Riemann-Lebesgue theorem, $\Lambda(p_1 \dots p_n) \to 0$ as $p_1^2+\dots+p_n^2 \to \infty$. Thus for any A and ϵ we can make $|A\Lambda| < \epsilon$ by choosing $p_1^2+\dots+p_n^2$ sufficiently large. Now, since L is L_1 and ψ_0 is bounded, the transform k_0 of ψ_0 is necessarily continuous, and, since I_i, $1 \leqslant i \leqslant N$, are closed, k_0 attains its upper and lower bounds in I_i. Therefore there is a δ so that $|k_0 - k_i| \leqslant c_i - \delta$ in I_i, $1 \leqslant i \leqslant N$. Also the upper bound of $|k_0|$ in $I_0 = c_0 - \delta_0$, $\delta_0 > 0$. If we make $|A\Lambda(p_1 \dots p_n)|$ less than $\min(\delta, \delta_0)$, $\psi_0 + A\sqrt{(2\pi)^{-n}}\, e^{i(p_1x_1+\dots+p_nx_n)}$ is also consistent with $\{k_i, c_i\}$.

In usual language the theorem states that ψ is not determined continuously by k if we define the distance between two (bounded) functions k and k_1 to be the least upper bound of $|k - k_1|$. This definition of distance will be used throughout the paper.

4. The preceding proof suggests that ψ is not determined by measurements of k because rapidly oscillating ψ have a small k. Accordingly, to get results about ψ from measurements of k, we should look for partial information about ψ which is insensitive to rapid oscillations of ψ, or else we must restrict the class $\mathscr{F}$ of ψ so as to exclude rapidly oscillating functions altogether.

In §§ II and III we obtain partial information about ψ, namely, suitably smoothed versions of ψ, and from them we get lower bounds for $\max|\psi|$. The results are applied to gravity survey.

In § IV we show for rather general kernels L that if ψ is continuous ψ itself is nearly determined by suitable smoothed versions of it constructed in § II, and therefore by measurements of k. A bound for the modulus of continuity of ψ may in practice be imposed in problems on the propagation of surface waves which are discussed at the end of the paper.

The result of § IV makes precise, and justifies, the remark at the beginning of the paragraph.

5. It is also shown that our result are complete in the sense that the results of §§ II to IV mentioned above are not only true for *any* ψ consistent with measurements $\{k_i, c_i\}$, but, under general conditions on L, for any $\{k_i, c_i\}$ there *is* some ψ which is consistent with $\{k_i, c_i\}$.

II. METHOD OF SMOOTHING ψ

6. In the present section we determine a smoothed version $\overline{\psi}$ of ψ, where

$$\overline{\psi}(x_1 \dots x_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \psi(u_1 \dots u_n)\, \mu(x_1 - u_1 \dots x_n - u_n)\, du_1 \dots du_n, \tag{6·1}$$

and $\mu(x_1 \dots x_n)$ is normalized so that

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mu(x_1 \dots x_n)\, dx_1 \dots dx_n = 1. \tag{6·2}$$

7. $\overline{\psi}$ may be expected to depend continuously on k if those (rapidly oscillating) ψ which have a small k also have a small $\overline{\psi}$.[†] For a given kernel L this can be secured by a suitable choice of μ. In particular, if L is rational $\mu = A\, e^{-\beta_1 x_1^2 - \dots - \beta_n x_n^2}$, $\beta > 0$, will do.

We carry out the formal statement and details in theorem A.

Theorem A

8. If we denote the Fourier transforms of L and μ by Λ, M, and if

(1) $\mu(x_1 \dots x_n)$, $L(x_1 \dots x_n)$ are L_1, (8·1)

(2) M/Λ is defined to be 0 if $M = 0$, and M/Λ is L_1, (8·2)

(3) $$\lambda(x_1 \dots x_n) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{M(p_1 \dots p_n)}{\Lambda(p_1 \dots p_n)}\, e^{i(p_1 x_1 + \dots + p_n x_n)}\, dp_1 \dots dp_n \text{ is } L_1, \tag{8·3}$$

(4) ψ is bounded and integrable in any finite interval; (8·4)

Then

(1) k is bounded and integrable in any finite interval, (8·5)

(2) $$\overline{\psi}(x_1 \dots x_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} k(\xi_1 \dots \xi_n)\, \lambda(x_1 - \xi_1 \dots x_n - \xi_n)\, d\xi_1 \dots d\xi_n, \tag{8·6}$$

(3) $\overline{\psi}$ depends continuously on k, that is if

$$\max |\, k(\xi_1 \dots \xi_n) - k_1(\xi_1 \dots \xi_n)\, | < \epsilon \quad \text{for} \quad (\xi_1 - x_1)^2 + \dots + (\xi_n - x_n)^2 < A,$$

then $$|\, \overline{\psi}(x_1 \dots x_n) - \overline{\psi}_1(x_1 \dots x_n)\, | < \delta(\epsilon, A),$$

where $$\delta(\epsilon, A) \to 0 \quad \text{when} \quad \epsilon \to 0 \quad \text{and} \quad A \to \infty. \tag{8·7}$$

[†] Why my reformulation may be useful appears from case (*a*) of the footnote on p. 160; instead of looking for the density distribution of sources ψ *at* a point x we look for the total of the sources in the neighbourhood of x. A practical application of the reformulation together with detailed numerical results is given in Bullard & Cooper (1948). Their paper is subsequently referred to as 'BC'.

Note also that, quite generally, for bounded f of L_1,

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} k(\xi_1\dots\xi_n)\, f(x_1-\xi_1\dots x_n-\xi_n)\, d\xi_1\dots d\xi_n$$

defines the smoothed version $\bar{\psi}$ with

$$\mu(x_1\dots x_n) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} L(x_1-u_1\dots x_n-u_n)\, f(u_1\dots u_n)\, du_1\dots du_n. \tag{8·8}$$

9. Note that the integral for λ exists since M/Λ is L_1 by (8·2). In para. 11 we state conditions on M/Λ which make λ L_1, as required by (8·3).

10. (i) We first prove (8·6) for functions ψ_A which are zero for large $x_1^2+\dots+x_n^2$. We define

$$\psi_A(x_1\dots x_n) = \psi(x_1\dots x_n) \quad \text{if} \quad x_1^2+\dots+x_n^2 < A,$$

$$\psi_A(x_1\dots x_n) = 0 \quad \text{if} \quad x_1^2+\dots+x_n^2 > A,$$

$$k_A(\xi_1\dots\xi_n) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \psi_A(x_1\dots x_n)\, L(\xi_1-x_1\dots\xi_n-x_n)\, dx_1\dots dx_n, \tag{10·1}$$

$$\bar{\psi}_A(x_1\dots x_n) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \psi_A(u_1\dots u_n)\, \mu(x_1-u_1\dots x_n-u_n)\, du_1\dots du_n. \tag{10·2}$$

Note that the range of integration of (10·1) and (10·2) is finite so that for any finite range ξ

$$\int_{\xi}\!\cdots\!\int |k_A|\, d\xi_1\dots d\xi_n \leqslant \int_{\xi}\!\cdots\!\int \int_{x_1^2+\dots+x_n^2<A}\!\!\!\cdots\!\int |\psi_A|\,|L(\xi_1-x_1\dots\xi_n-x_n)|\, d\xi_1\dots d\xi_n\, dx_1\dots dx_n$$

$$\leqslant \left(\int_{x_1^2+\dots+x_n^2<A}\!\!\!\cdots\!\int |\psi_A|\, dx_1\dots dx_n\right)\left(\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} |L(u_1\dots u_n)|\, du_1\dots du_n\right).$$

Since L is L_1, and $|\psi_A|$ is bounded, k_A is L_1; and since μ is L_1, $\bar{\psi}_A$ is also L_1.

We may take Fourier transforms of (10·1) and (10·2) and get

$$\frac{1}{\sqrt{(2\pi)^n}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} k_A\, e^{i(p_1\xi_1+\dots+p_n\xi_n)}\, d\xi_1\dots d\xi_n$$

$$= \frac{1}{\sqrt{(2\pi)^n}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \int_{x_1^2+\dots+x_n^2<A}\!\!\!\cdots\!\int \psi_A\, L(\xi_1-x_1\dots\xi_n-x_n)\, e^{i(p_1\xi_1+\dots+p_n\xi_n)} \times d\xi_1\dots d\xi_n\, dx_1\dots dx_n$$

$$= \Lambda(p_1\dots p_n) \int_{x_1^2+\dots+x_n^2<A}\!\!\!\cdots\!\int \psi_A\, e^{i(p_1x_1+\dots+p_nx_n)}\, dx_1\dots dx_n,$$

where the order of integration may be inverted by absolute convergence of the integrals. Hence, if we denote the Fourier transforms of ψ_A, k_A, $\bar{\psi}_A$ by Ψ_A, K_A, $\bar{\Psi}_A$,

$$K_A = \sqrt{(2\pi)^n}\, \Psi_A \Lambda, \tag{10·3}$$

and similarly

$$\bar{\Psi}_A = \sqrt{(2\pi)^n}\, \Psi_A M, \tag{10·4}$$

where $\overline{\Psi}_A = 0$ when $M = 0$. By (10·3) and (10·4)

$$\overline{\Psi}_A = (M/\Lambda)\,K_A.$$

Since k_A and M/Λ are L_1, applying the Faltungstheorem (Titchmarsh 1937, theorem 35), we get

$$\overline{\psi}_A(x_1 \dots x_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} k_A(\xi_1 \dots \xi_n)\,\lambda(x_1-\xi_1 \dots x_n-\xi_n)\,d\xi_1 \dots d\xi_n. \tag{10·5}$$

We have expressed $\overline{\psi}_A$ as a (continuous) linear transform of k_A.

(ii) We may drop the restriction that $\psi(x_1 \dots x_n) = 0$ for $x_1^2 + \dots + x_n^2 > A$.

Without loss of generality we consider $\overline{\psi}(0 \dots 0)$, which is obtained from $\overline{\psi}(x_1 \dots x_n)$ by a shift of origin.

Observe that

$$|\,k\,| < \max|\,\psi\,| \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |\,L(u_1 \dots u_n)\,|\,du_1 \dots du_n;$$

since $|\,\psi\,|$ is bounded and L is L_1, k is bounded. Also it is integrable in any finite interval.

(iii) We estimate

$$\left|\overline{\psi}(0 \dots 0) - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} k(\xi_1 \dots \xi_n)\,\lambda(-\xi_1 \dots -\xi_n)\,d\xi_1 \dots d\xi_n\right|,$$

where the integral converges since k is bounded and integrable by (ii), and λ is L_1 by (8·3).

(*a*) Since μ is L_1 and ψ is bounded ($<c$), for $A > A_0(\epsilon)$

$$\begin{aligned}\overline{\psi}(0 \dots 0) - \overline{\psi}_A(0 \dots 0)\,| &= \left|\ \underset{x_1^2+\dots+x_n^2>A}{\int \dots \int} \psi(x_1 \dots x_n)\,\mu(-x_1 \dots -x_n)\,dx_1 \dots dx_n\right| \\ &\leqslant c \underset{x_1^2+\dots+x_n^2>A}{\int \dots \int} |\,\mu(-x_1 \dots -x_n)\,|\,dx_1 \dots dx_n < \tfrac{1}{3}\epsilon.\end{aligned} \tag{10·6}$$

(*b*) Since L is L_1, for $\xi_1^2 + \dots + \xi_n^2 < \frac{1}{4}A \quad (A > A_1(\epsilon))$,

$$\begin{aligned}|\,k(\xi_1 \dots \xi_n) - k_A(\xi_1 \dots \xi_n)\,| &= \left|\ \underset{x_1^2+\dots+x_n^2>A}{\int \dots \int} \psi(x_1 \dots x_n)\,L(\xi_1-x_1 \dots \xi_n-x_n)\,dx_1 \dots dx_n\right| \\ &< c \underset{u_1^2+\dots+u_n^2>\frac{1}{4}A}{\int \dots \int} |\,L(u_1 \dots u_n)\,|\,du_1 \dots du_n < \tfrac{1}{3}\epsilon/C_0,\end{aligned} \tag{10·7}$$

where $$C_0 = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |\,\lambda(x_1 \dots x_n)\,|\,dx_1 \dots dx_n \quad (<\infty \text{ by } (8·3)).$$

Also, for all $(\xi_1 \dots \xi_n)$

$$|\,k\,| \leqslant c \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |\,L(u_1 \dots u_n)\,|\,du_1 \dots du_n = cC_1, \text{ say,} \tag{10·8}$$

and $|\,k_A\,| \leqslant cC_1$.

(*c*) Since λ is L_1, for $A > A_2(\epsilon)$

$$2cC_1 \underset{x_1^2+\dots+x_n^2>\frac{1}{4}A}{\int\cdots\int} |\,\lambda(x_1\dots x_n)\,|\,dx_1\dots dx_n < \tfrac{1}{3}\epsilon. \tag{10·9}$$

Now choose $A > \max(A_0, A_1, A_2)$:

$$\begin{aligned}\left|\overline{\psi}(0\dots0)-\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}k(\xi_1\dots\xi_n)\,\lambda(-\xi_1\dots-\xi_n)\,d\xi_1\dots d\xi_n\right| &\leqslant |\,\overline{\psi}(0\dots0)-\overline{\psi}_A(0\dots0)\,| \\ &+\left|\overline{\psi}_A(0\dots0)-\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}k_A(\xi_1\dots\xi_n)\,\lambda(-\xi_1\dots-\xi_n)\,d\xi_1\dots d\xi_n\right| \\ &+\left|\underset{\xi_1^2+\dots+\xi_n^2<\frac{1}{4}A}{\int\cdots\int}[k_A(\xi_1\dots\xi_n)-k(\xi_1\dots\xi_n)]\,\lambda(-\xi_1\dots-\xi_n)\,d\xi_1\dots d\xi_n\right| \\ &+\left|\underset{\xi_1^2+\dots+\xi_n^2>\frac{1}{4}A}{\int\cdots\int}[k_A(\xi_1\dots\xi_n)-k(\xi_1\dots\xi_n)]\,\lambda(-\xi_1\dots-\xi_n)\,d\xi_1\dots d\xi_n\right| < \epsilon\end{aligned}$$

by (10·6), (10·5), (10·7) and (10·9).

Since ϵ is arbitrary this proves (8·6).

(iv) To prove (8·7):

$$\begin{aligned}&|\,\overline{\psi}(x_1\dots x_n)-\overline{\psi}_1(x_1\dots x_n)\,| \\ &\leqslant\left|\underset{(\xi_1-x_1)^2+\dots+(\xi_n-x_n)^2<A}{\int\cdots\int}[k(\xi_1\dots\xi_n)-k_1(\xi_1\dots\xi_n)]\,\lambda(x_1-\xi_1\dots x_n-\xi_n)\,d\xi_1\dots d\xi_n\right| \\ &\quad+\left|\underset{(\xi_1-x_1)^2+\dots+(\xi_n-x_n)^2>A}{\int\cdots\int}[k(\xi_1\dots\xi_n)-k_1(\xi_1\dots\xi_n)]\,\lambda(x_1-\xi_1\dots x_n-\xi_n)\,d\xi_1\dots d\xi_n\right| \\ &\leqslant\epsilon\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}|\lambda(u_1\dots u_n)\,|\,du_1\dots du_n \\ &\quad+\max_{(\xi_1-x_1)^2+\dots+(\xi_n-x_n)^2>A}|\,k-k_1\,|\underset{u_1^2+\dots+u_n^2>A}{\int\cdots\int}|\,\lambda(u_1\dots u_n)\,|\,du_1\dots du_n.\end{aligned}$$

Hence (8·7).

11. Necessary and sufficient conditions are known which ensure that the Fourier transform of a function is L_1 (Doss 1948, p. 173).† As a matter of fact, it is not easy to verify for practical M/Λ if these conditions hold, and in practice one would use some simple conditions as in Titchmarsh (1937, 6·13); a slight modification of them covers the functions M/Λ used in the examples of paras. 13 to 15.

12. It now remains to show that $\overline{\psi}(x)$ is not only determined by $\{k_i, c_i\}$, but that there exists a ψ which is consistent with $\{k_i, c_i\}$. We prove that, for suitable μ, $\overline{\psi}$ itself is such a ψ.

---

† I am grateful to the referee for calling my attention to this paper, and for his many detailed comments.

If theorem A holds for sufficiently steep μ, that is, if for any ϵ there is a μ so that

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(|x_1|+\ldots+|x_n|)\,|\mu(x_1\ldots x_n)|\,dx_1\ldots dx_n<\epsilon, \tag{12·1}$$

then for any $\{k_i, c_i\}$ there is a ψ which is consistent with $\{k_i, c_i\}$. (12·2)

We prove the result for $n = 1$; the argument can also be carried through for several independent variables, but the details are much more complicated. Observe first, by integrating (1·1), that the transform of $\overline{\psi}$ is

$$\bar{k}(\xi) = \int_{-\infty}^{\infty} k(x)\,\mu(\xi-x)\,dx. \tag{12·3}$$

We now choose a $k_1(\xi)$ consistent with $\{k_i, c_i\}$ and show that under (12·1) $\bar{k}_1(\xi)$, the transform of $\overline{\psi}_1(x)$, is also consistent with $\{k_i, c_i\}$. By (12·3) $\overline{\psi}_1(x)$ is a ψ consistent with $\{k_i, c_i\}$, as required.

Definition of $k_1(\xi)$

Suppose the intervals I_i, I_{i+1} abut along the line $\xi = \xi'_i$, and I_i is to the left of I_{i+1}; then

$$\max(k_i-c_i, k_{i+1}-c_{i+1}) < k(\xi'_i) < \min(k_i+c_i, k_{i+1}+c_{i+1}) \quad (1 \leqslant i < N).$$

At the left-hand end ξ'_0 of I_1

$$\max(-c_0, k_1-c_1) < k(\xi'_0) < \min(c_0, k_1+c_1),$$

and at the right-hand end ξ'_N of I_N

$$\max(-c_0, k_N-c_N) < k(\xi'_N) < \min(c_0, k_N+c_N).$$

Denote

$$\left.\begin{array}{lll} \tfrac{1}{2}\{\max(k_i-c_i, k_{i+1}-c_{i+1})+\min(k_i+c_i, k_{i+1}+c_{i+1})\} & \text{by} & \eta'_i \\ \tfrac{1}{2}\{\max(-c_0, k_1-c_1)+\min(c_0, k_1+c_1)\} & \text{by} & \eta'_0 \\ \tfrac{1}{2}\{\max(-c_0, k_N-c_N)+\min(c_0, k_N+c_N)\} & \text{by} & \eta'_N \end{array}\right\} \quad (1 \leqslant i < N).$$

Define $k_1(\xi)$ as follows: for

$$\xi \leqslant \xi'_0 \quad k_1(\xi) = \eta'_0;$$

$$\xi \geqslant \xi'_N \quad k_1(\xi) = \eta'_N;$$

$$\xi'_0 < \xi < \xi'_N\text{: join } (\xi'_i, \eta'_i) \text{ to } (\xi'_{i+1}, \eta'_{i+1}), \quad 1 \leqslant i < N,$$

by a straight line. Observe that $k_1(\xi)$ is everywhere at a distance $\geqslant \delta$, say, from the bounds (3·1), (3·2), and the modulus of its gradient nowhere exceeds

$$\max|\eta'_{i+1}-\eta'_i|/(\xi'_{i+1}-\xi'_i), \quad 0 \leqslant i < N, \quad = \gamma, \text{ say.}$$

Now choose $\epsilon = \tfrac{1}{2}(\delta/\gamma)$ in (12·1):

$$|\bar{k}_1(\xi)-k_1(\xi)| = \left|\int_{-\infty}^{\infty}[k_1(x)-k_1(\xi)]\,\mu(\xi-x)\,dx\right|$$

$$<\gamma\int_{-\infty}^{\infty}|\xi-x|\,|\mu(\xi-x)|\,dx = \gamma\int_{-\infty}^{\infty}|x|\,|\mu(x)|\,dx<\tfrac{1}{2}\delta$$

by (6·2) and (12·1).

Thus also $\bar{k}_1$, and therefore $\bar{\psi}_1$, is consistent with $\{k_i, c_i\}$. Observe for reference that the condition (12·1) is satisfied by $\mu(x) = \sqrt{(\beta/\pi)}\,\mathrm{e}^{-\beta x^2}$, if $\beta > \pi^{-1}\epsilon^{-2}$.

An approximate expression for $\bar{\psi}(x)$ which is convenient for applications is given by

$$\bar{\psi}_M(x) = \sum_1^N k_i \int_{I_i} \lambda(x-\xi)\,d\xi,$$

where, by theorem A, for any $\psi(x)$ consistent with $\{k_i, c_i\}$

$$|\,\bar{\psi}(x) - \bar{\psi}_M(x)\,| < \sum_0^N c_i \int_{I_i} |\,\lambda(x-\xi)\,|\,d\xi.$$

$\bar{\psi}_M(x)$ is calculated by 'BC' for particular λ, and discussed there in detail.

Examples of theorem A

13. We now apply theorem A to the gravitational problem of 'BC'.

The work introduces the transformation of integrals needed in two-dimensional and three-dimensional problems, the latter giving rise to Bessel functions. The integrals $\lambda(x)$, $\lambda(x,y)$ corresponding to the smoothing functions $\mathrm{e}^{-\beta x^2}$, $\mathrm{e}^{-\beta(x^2+y^2)}$ are tabulated in 'BC'.

14. The normal field $k(\xi)$ on a plane Σ due to a one-dimensional distribution of masses of line density $\psi(x)$ on a plane Σ' at a distance α from Σ satisfies

$$k(\xi) = 2G\alpha \int_{-\infty}^{\infty} \frac{\psi(x)\,dx}{(\xi-x)^2+\alpha^2}, \tag{14·1}$$

where G is the gravitational constant.

We suppose $k(\xi)$, $\psi(x)$ to be bounded, and integrable in any finite interval. If we choose as our smoothing function

$$\mu(x) = \sqrt{\frac{\beta}{\pi}}\,\mathrm{e}^{-\beta x^2} \quad \text{so that} \quad \bar{\psi}(x) = \sqrt{\frac{\beta}{\pi}} \int_{-\infty}^{\infty} \psi(u)\,\mathrm{e}^{-\beta(x-u)^2}\,du,$$

then

$$\bar{\psi}(x) = \int_{-\infty}^{\infty} k(\xi)\,\lambda(x-\xi)\,d\xi, \quad \text{where} \quad \lambda(x) = \frac{1}{(2\pi)^2\,G} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{4}p^2/\beta+\alpha|p|}\,\mathrm{e}^{ixp}\,dp.$$

Observe that

$$M(p) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} \mu(x)\,\mathrm{e}^{ipx}\,dx = \frac{\mathrm{e}^{-\frac{1}{4}p^2/\beta}}{\sqrt{(2\pi)}},$$

and

$$\Lambda(p) = \frac{2\alpha G}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} \frac{\mathrm{e}^{ipx}}{x^2+\alpha^2}\,dx = \sqrt{(2\pi)}\,G\,\mathrm{e}^{-\alpha|p|}$$

by contour integration round a semicircle in the upper half-plane of x for $p > 0$, in the lower half-plane for $p < 0$. Thus $M/\Lambda = (2\pi G)^{-1}\,\mathrm{e}^{-\frac{1}{4}p^2/\beta+\alpha|p|}$, which is L_1.

It is easily verified that λ is also L_1 by integrating

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{4}p^2/\beta+\alpha|p|}\,\mathrm{e}^{ipx}\,dp = 2\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{4}p^2/\beta+\alpha p} \cos px\,dp$$

twice by parts, and showing that $\lambda(x) = O(|\,x\,|^{-2})$.

Hence by theorem A the results follow.

15. The normal field $k(\xi,\eta)$ on Σ due to a two-dimensional distribution of masses of surface density $\psi(x,y)$ on Σ' satisfies

$$k(\xi,\eta) = G\alpha\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\psi(x,y)\,dx\,dy}{[(\xi-x)^2+(\eta-y)^2+\alpha^2]^{\frac{3}{2}}}$$

If we choose as smoothing function

$$\mu(x,y) = \mu(x)\,\mu(y), \quad \text{so that} \quad \overline{\psi}(x,y) = \frac{\beta}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi(u,v)\,\mathrm{e}^{-\beta[(x-u)^2+(y-v)^2]}\,du\,dv,$$

$$\overline{\psi}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}k(\xi,\eta)\,\lambda(x-\xi,y-\eta)\,d\xi\,d\eta,$$

where
$$\lambda(x,y) = \frac{1}{(2\pi)^2\,G}\int_0^{\infty}\rho\,\mathrm{e}^{-\frac{1}{4}\rho^2/\beta+\alpha\rho}\,J_0(\rho\sqrt{[x^2+y^2]})\,d\rho.$$

(i) $M(p_1,p_2) = \dfrac{1}{2\pi}\mathrm{e}^{-\frac{1}{4}(p_1^2+p_2^2)/\beta}$ as in para. 14.

(ii) $$\Lambda(p_1p_2) = \frac{\alpha G}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\mathrm{e}^{i(p_1x_1+p_2y)}\,dx\,dy}{(x^2+y^2+\alpha^2)^{\frac{3}{2}}} = \alpha G\int_0^{\infty}\frac{rJ_0[r\sqrt{(p_1^2+p_2^2)}]\,dr}{(r^2+\alpha^2)^{\frac{3}{2}}}.$$

To see this, rotate the axes so that

$$p_1x+p_2y = x'\sqrt{(p_1^2+p_2^2)}, \quad -p_2x+p_1y = y'\sqrt{(p_1^2+p_2^2)}.$$

Then $x^2+y^2 = x'^2+y'^2$, and, changing to polar co-ordinates, we get

$$\Lambda(p_1,p_2) = \frac{\alpha G}{2\pi}\int_0^{\infty}r\,dr\int_0^{2\pi}\frac{\mathrm{e}^{-ir\cos\theta\sqrt{(p_1^2+p_2^2)}}\,d\theta}{(r^2+\alpha^2)^{\frac{3}{2}}},$$

independent of the order of integration by absolute convergence of the integral.

Since
$$\int_0^{2\pi}\mathrm{e}^{iPr\cos\theta}\,d\theta = 2\pi J_0(rP),$$
we get the result.

(iii) By Titchmarsh (1937, equation (7·4·8))

$$\int_0^{\infty}x\,\mathrm{e}^{-\alpha x}\,J_0(ux)\,dx = \frac{\alpha}{(\alpha^2+u^2)^{\frac{3}{2}}},$$

so that by Hankel's theorem

$$\Lambda(p_1,p_2) = G\,\mathrm{e}^{-\alpha\sqrt{(p_1^2+p_2^2)}} \quad (\text{put } f(y) = \sqrt{y}\,\mathrm{e}^{-\alpha y},\ x = \sqrt{(p_1^2+p_2^2)}$$

in equation (8·18·1) of Titchmarsh (1937).

By (8·3)
$$\lambda(x,y) = \frac{1}{(2\pi)^3\,G}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{4}(p_1^2+p_2^2)/\beta+\alpha\sqrt{(p_1^2+p_2^2)}}\,\mathrm{e}^{i(p_1x+p_2y)}\,dp_1\,dp_2.$$

By a rotation of axes (p_1,p_2) similar to (ii) we get the required expression for $\lambda(x,y)$.

To verify that λ is L_1 we show that $\lambda = O(x^2+y^2)^{-\frac{3}{2}}$. Integrate

$$\int_0^\infty \rho\, \mathrm{e}^{-\frac{1}{4}\rho^2/\beta+\alpha\rho}\, J_0(\rho r)\, d\rho$$

three times by parts, using the relations

$$r\rho J_0(\rho r) = \frac{d}{d\rho}\rho J_1(\rho r), \quad rJ_1(\rho r) = -\frac{dJ_0(\rho r)}{d\rho}.$$

III. A LOWER BOUND FOR $\max|\psi|$

16. In the present section we determine a lower bound for $\max|\psi|$. Such a bound is certainly determined by measurements $\{k_i, c_i\}$, since the maximum of $|\psi|$ is not less than that of the smoothed versions of ψ obtained in theorem A, which are known to be determined by $\{k_i, c_i\}$. In any case

$$\max|\psi| \geqslant \frac{\max|k|}{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}|L(u_1\ldots u_n)|\,du_1\ldots du_n}. \tag{16·1}$$

Now, by para. 3, rapidly oscillating ψ have a small k, and it is easy to see that smooth ψ have a smooth k. Thus if k oscillates rapidly with noticeable amplitude the lower bound for $\max|\psi|$ should be large. This is got by a modification of (16·1). Suppose there are two points $(\xi_1\ldots\xi_n)$, $(\xi_1+\epsilon_1\ldots\xi_n+\epsilon_n)$, where any k consistent with $\{k_i, c_i\}$ differs by at least ω. Then

$$\begin{aligned}\omega &\leqslant |k(\xi_1+\epsilon_1\ldots\xi_n+\epsilon_n)-k(\xi_1\ldots\xi_n)|\\ &\leqslant \max|\psi|\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}|L(u_1+\epsilon_1\ldots u_n+\epsilon_n)-L(u_1\ldots u_n)|\,du_1\ldots du_n,\end{aligned} \tag{16·2}$$

where the integral is small if the ϵ's are small.

Also in a variety of applications where the kernel depends on a parameter α, oscillating ψ have a small k for large α, i.e. $\Lambda(p_1\ldots p_n;\alpha)\to 0$ as $\alpha\to\infty$ unless $|p_1|+\ldots+|p_n| = 0$. Then for a given oscillatory k the lower bound for $\max|\psi|$ should be large if α is large. To obtain such a bound consider an arbitrary smoothed version of ψ (8·8), defined by

$$\begin{aligned}&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}k(\xi_1\ldots\xi_n)\,\lambda(x_1-\xi_1\ldots x_n-\xi_n)\,d\xi_1\ldots d\xi_n\\ &\quad=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\psi(u_1\ldots u_n)\left(\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}L(\xi_1-u_1\ldots\xi_n-u_n)\right.\\ &\qquad\qquad\left.\times\,\lambda(x_1-\xi_1\ldots x_n-\xi_n)\,d\xi_1\ldots d\xi_n\right)du_1\ldots du_n\\ &\quad\leqslant\max|\psi|\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left|\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}L(\xi_1-u_1\ldots\xi_n-u_n)\right.\\ &\qquad\qquad\left.\times\,\lambda(x_1-\xi_1\ldots x_n-\xi_n)\,d\xi_1\ldots d\xi_n\right|du_1\ldots du_n.\end{aligned} \tag{16·3}$$

What is wanted is a λ so that for some $(x_1 \dots x_n)$

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} k(\xi_1 \dots \xi_n)\,\lambda(x_1-\xi_1 \dots x_n-\xi_n)\,d\xi_1 \dots d\xi_n \geqslant \omega > 0,$$

provided only that k is consistent with $\{k_i, c_i\}$, and

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left| \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L(\xi_1-u_1 \dots \xi_n-u_n)\,\lambda(x_1-\xi_1 \dots x_n-\xi_n)\,d\xi_1 \dots d\xi_n \right| \times du_1 \dots du_n \to 0 \quad \text{as} \quad \alpha \to \infty.$$

We consider functions of one variable, $n = 1$, write $\lambda(\xi) = f(\xi)\,e^{ip\xi}$, shift the origin to make $x_1 = 0$, and for given $\{k_i, c_i\}$ we construct a suitable f in

Theorem B

17. (1) Suppose $k(\xi)$ and $\psi(x)$ satisfy the relation (1·1); $L(\zeta;\,\alpha)$ is a rational function of ζ ($\zeta = \xi + i\eta$) for each α, real and L_1 on the real axis, with poles $\xi_n(\alpha) \pm i\eta_n(\alpha)$ and residues $r_n(\alpha)$ at

$$\zeta_n(\alpha), \quad 1 \leqslant n \leqslant N_1; \tag{17·1}$$

further, there is a pair i, j, $i \neq 0$, $j \neq 0$, so that

$$k_i - c_i \geqslant k_j + c_j + \gamma, \quad \gamma > 0; \tag{17·2}$$

that is, no constant k satisfies $|k - k_i| < c_i$.

Then:

There is a rational function $f(\zeta)$ regular in the upper half-plane, and L_1 on the real axis, a $p > 0$ and an $\omega > 0$, so that

$$\left\{\sum_1^{N_1} e^{-p\eta_n(\alpha)}\,|r_n(\alpha)| \int_{-\infty}^{\infty} |f[\xi + i\eta_n(\alpha)]|\,d\xi\right\}^{-1} \leqslant \frac{2\pi \max|\psi|}{\left|\int_{-\infty}^{\infty} k(\xi)\,f(\xi)\,e^{ip\xi}\,d\xi\right|}, \tag{17·3}$$

and for any $k(\xi)$ consistent with $\{k_i c_i\}$

$$\left|\int_{-\infty}^{\infty} k(\xi)\,f(\xi)\,e^{ip\xi}\,d\xi\right| > \omega. \tag{17·4}$$

The details of the construction of $f(\zeta)$, p and ω are summarized at the end of the proof.

(2) *Inequalities for* α. If $\max|\psi|$ is prescribed, and if for $p > 0$,

$$-\log|r_n(\alpha)| + p\eta_n(\alpha) \to \infty$$

as $\alpha \to \infty$, (17·3) gives a bound for α; the right-hand side is independent of α, and finite for some p; since $\int_{-\infty}^{\infty} |f[\xi + i\eta_n(\alpha)]|\,d\xi$ is uniformly bounded for $\eta_n(\alpha) > 0$, the left-hand side $\to \infty$ as $\alpha \to \infty$ so that α must not be too large.[†] (17·5)

[†] It is easily shown that under the conditions of theorem B the integral of $|\partial L/\partial \xi| \to 0$ as $\alpha \to \infty$, so that (16·2) provides a bound for α; for instance in the example of para. 22 (16·2) yields a satisfactory bound. By introducing the function $f(\zeta)$ we can often give a bound for α when (16·2) cannot be applied at all because $\partial L/\partial \xi$ is not L_1, or we get a better one; if $\psi(x, y) = 0$ for $y > \beta$, β large and fixed, (16·2) can be applied to the example of para. 23, but the bound is poor.

18. Lemma I. If $f(\zeta)$ is a rational function, regular in the upper half-plane, and L_1 on the real axis, $L(\zeta;\, \alpha)$ satisfies (17·1), $\psi(x)$ is bounded and integrable in any finite interval, and $p>0$, then (17·3) and (17·5) hold.

Assume

$$\int_{-\infty}^{\infty} k(\xi)\, f(\xi)\, e^{ip\xi}\, d\xi \neq 0,$$

since otherwise there is nothing to prove.

Observe that $f(\zeta)$ and $L(\zeta)$ are $O(|\zeta|^{-2})$ for large $|\zeta|$, since they are rational and L_1 on the real axis. Therefore

$$\begin{aligned}\int_{-\infty}^{\infty} k(\xi)\, f(\xi)\, e^{ip\xi}\, d\xi &= \int_{-\infty}^{\infty} f(\xi)\, e^{ip\xi}\left(\int_{-\infty}^{\infty} \psi(x)\, L(\xi-x;\, \alpha)\, dx\right) d\xi \\ &= \int_{-\infty}^{\infty} \psi(x)\left(\int_{-\infty}^{\infty} f(\xi)\, L(\xi-x;\, \alpha)\, e^{ip\xi}\, d\xi\right) dx,\end{aligned}$$

where the inversion of the order of integration is justified, since the integrals are absolutely convergent.

We integrate $f(\zeta)\, L(\zeta-x;\, \alpha)\, e^{ip\zeta}$ round the contour of a large semicircle in the upper half-plane; since $p>0$, the integrand is $O(|\zeta|^{-4})$ over the circular arc; we may apply the calculus of residues, and find

$$\int_{-\infty}^{\infty} f(\xi)\, L(\xi-x;\, \alpha)\, e^{ip\xi}\, d\xi = 2\pi i \sum_{1}^{N_1} f[x+\xi_n(\alpha)+i\eta_n(\alpha)]\, r_n(\alpha)\, e^{ipx}\, e^{ip\xi_n(\alpha)}\, e^{-p\eta_n(\alpha)}.$$

Also $\int_{-\infty}^{\infty} |f[x+i\eta_n(\alpha)]|\, dx$ is uniformly bounded in $\eta_n(\alpha)>0$, since

$$|f(x+iy)| = O\,|x^2+y^2|^{-1},$$

so that the result (17·3) holds.

The inequality for α obtained by (17·5) also holds, since the argument of (17·5) only requires the equation (17·3) to be true.

Note that if $f(\zeta)$ had singularities in the upper half-plane we could not conclude that $\int_{-\infty}^{\infty} |f[x+i\eta_n(\alpha)]|\, dx<\infty$. Also, if it were only assumed that $\int_{-\infty}^{\infty} k(\xi)\, f(\xi)\, e^{ip\xi}\, d\xi \neq 0$ for some p, p might be <0, and the argument of the lemma would break down.

19. We now have to construct a function $f(\zeta)$ which ensures that for some $p>0$ (17·4) holds.

In lemma II we state sufficient conditions on $f(\zeta)$, and in lemma III we define a function satisfying these conditions.

Lemma II. Suppose (1) $k_i-c_i \geqslant k_j+c_j+\gamma$, or $k_i+c_i+\gamma \leqslant k_j-c_j$, $\gamma>0$, I_i, I_j are equal and symmetrical about the origin, I_j to the right of I_i; (19·1)

(2) $f(\zeta)$ is rational, regular in the upper half plane, and L_1 on the real axis; (19·2)

(3) $\Re[f(\xi)]$ is antisymmetrical, and $\geqslant 0$ on I_i, (19·3)

$$(4)\ \int_{J_{ij}} |\Re[f(\xi)]|\,d\xi < \frac{\gamma}{4\pi \max\limits_{0\leqslant i\leqslant N}(|k_i|+c_i)} \int_{I_i} \Re[f(\xi)]\,d\xi; \tag{19·4}$$

where J_{ij} is the complement of I_i+I_j.

Then
$$\left|\int_{-\infty}^{\infty} k(\xi)\,f(\xi)\,e^{ip\xi}\,d\xi\right| > \frac{\gamma}{8}\int_{I_i} \Re[f(\xi)]\,d\xi > 0, \tag{19·5}$$

provided p is so small that

$$\sum_{i=0}^{N}(|k_i|+c_i)\int_{I_i} |\Im[f(\xi)]|\,|\sin p\xi|\,d\xi < \frac{\gamma}{8}\int_{I_i} \Re[f(\xi)]\,d\xi. \tag{19·6}$$

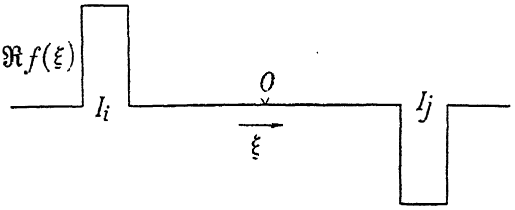


FIGURE 1

(i) Observe first that if in unequal intervals

$$I_i',\quad I_j',\quad k_i-c_i \geqslant k_j+c_j+\gamma \quad\text{or}\quad k_i+c_i+\gamma \leqslant k_j-c_j$$

we take equal subintervals I_i, I_j or I_i', I_j' and make them symmetrical about the origin by a shift of origin; I_i, I_j then satisfy (19·1).

(ii) Observe also that (19·5) always has a solution: since $f(\xi)$ is L_1 we can make

$$\int_X^{\infty} |\Im[f(\xi)]|\,|\sin p\xi|\,d\xi < \int_X^{\infty} |\Im[f(\xi)]|\,d\xi,$$

small by choosing X large; then $\int_{-X}^{X} |\Im[f(\xi)]|\,|\sin p\xi|\,d\xi$ is made small by choosing p small.

(iii) If p is so small that the distance between the outermost end-points of I_i and I_j is less than $2\pi/(3p)$, $\cos p\xi > \frac{1}{2}$ when ξ lies in I_i or I_j, and, by (19·1),

$$\left|\int_{I_i+I_j} k(\xi)\,\Re[f(\xi)]\cos p\xi\,d\xi\right| > \tfrac{1}{2}\gamma\int_{I_i} \Re[f(\xi)]\,d\xi$$

for any k consistent with $\{k_i, c_i\}$.

Also
$$\left|\int_{J_{ij}} k(\xi)\,\Re[f(\xi)]\cos p\xi\,d\xi\right| < \int_{J_{ij}} |k(\xi)\,\Re[f(\xi)]|\,d\xi < \tfrac{1}{4}\gamma\int_{I_i} \Re[f(\xi)]\,d\xi$$

by (20·3), and therefore

$$\left|\int_{-\infty}^{\infty} k(\xi)\,\Re[f(\xi)]\cos p\xi\,d\xi\right| > \tfrac{1}{4}\gamma\int_{I_i} \Re[f(\xi)]\,d\xi. \tag{19·7}$$

But $$\left|\int_{-\infty}^{\infty} k(\xi)\, f(\xi)\, e^{ip\xi}\, d\xi\right| \geqslant \left|\Re \int_{-\infty}^{\infty} k(\xi)\, f(\xi)\, e^{ip\xi}\, d\xi\right|$$
$$= \left|\int_{-\infty}^{\infty} k(\xi)\, \Re[f(\xi)] \cos p\xi\, d\xi - \int_{-\infty}^{\infty} k(\xi)\, \Im[f(\xi)] \sin p\xi\, d\xi\right|.$$

By (19·7) and (19·6) the lemma follows.

20. Lemma III. For any $\gamma > 0$ and $\max\limits_{1 \leqslant i \leqslant N} (|k_i| + c_i) > 0$ we can construct a function satisfying (19·2), (19·3) and (19·4).

(*a*) Map the upper half-plane $\eta > 0$ on the unit circle $|z| < 1$ so that $\zeta = 0$ goes into $z = -i$, the point at infinity goes into $z = i$, and the interval I_j goes into the arc $-\delta < \arg z < \delta$, $|z| = 1$, $\delta < \frac{1}{2}\pi$.

The transformation function is $$z = i\frac{\zeta - ia}{\zeta + ia},$$

where $$e^{-i\delta} = i\frac{\rho - ia}{\rho + ia}, \quad e^{i\delta} = i\frac{\rho_1 - ia}{\rho_1 + ia} \quad \text{and} \quad (\rho, \rho_1)$$

are the end-points of I_j. Thus a and δ are obtained in terms of the length of I_j and the distance between I_i and I_j.

Note that $$a > 0, \quad \text{since} \quad \delta < \tfrac{1}{2}\pi, \tag{20·1}$$

points symmetric about the η-axis are mapped into points symmetric about the y-axis, (20·2)

the ξ-axis goes into the circle $z = e^{i\theta}$ $(0 \leqslant \theta < 2\pi)$, and we define the transformation $\theta(\xi)$ by $$e^{i\theta} = i\frac{\xi - ia}{\xi + ia}. \tag{20·3}$$

(*b*) Let $$\mathfrak{F}_n(\theta) = \frac{(1 - \cos\theta)^n - (1 + \cos\theta)^n}{\int_{-\pi}^{\pi} (1 + \cos\theta)^n\, d\theta} \cos^2\theta.$$

Then the (unique) regular function in $\eta > 0$ which satisfies
$$\Re[f_n(\xi)] = \mathfrak{F}_n[\theta(\xi)], \tag{20·4}$$
$$f_n(\xi) \to 0 \quad \text{as} \quad \xi \to \infty \tag{20·5}$$
is rational and L_1 on the real axis, (20·6)
and satisfies (19·3).

(i) Denote $$f_n[\zeta(z)] \quad \text{by} \quad g_n(z).$$

If $f_n(\zeta)$ is regular in $\eta > 0$, and satisfies (20·4) and (20·5), $g_n(z)$ is regular in $|z| < 1$, $\Re[g_n(e^{i\theta})] = \mathfrak{F}_n(\theta)$ on $z = e^{i\theta}$ $(0 \leqslant \theta < 2\pi)$, $g_n(i) = 0$.

$y_n(z)$ is unique; for, since $\mathfrak{F}_n(\theta)$ is continuous there is a unique harmonic function in $|z| < 1$ with boundary values $\mathfrak{F}_n(\theta)$; its conjugate is determined except for an arbitrary constant which is fixed by the condition $g_n(i) = 0$.

(ii) $g_n(z)$ is a polynomial.

Let m be the integral part of $\frac{1}{2}(n - 1)$. Then
$$(1 + \cos\theta)^n - (1 - \cos\theta)^n = 2\sum_{r=0}^{m} {}_nC_{2r+1} \cos^{2r+1}\theta = \sum_{r=0}^{m} a_r \cos(2r + 1)\theta,$$

where a_r are real since an odd power $\cos^{2r+1}\theta$ can be expressed as a sum of terms $\cos(2s+1)\theta$, $0 \leqslant s \leqslant r$, with real coefficients.

Observe that

$$2\cos^2\theta = \cos 2\theta + 1 \quad \text{and} \quad 2\cos(2r+1)\theta\cos 2\theta = \cos(2r+3)\theta + \cos(2r-1)\theta,$$

so that $\mathfrak{F}_n(\theta) = A\{(1+\cos\theta)^n - (1-\cos\theta)^n\}\cos^2\theta = \sum_0^{m+1} b_r \cos(2r+1)\theta$ with real b

$$= \Re\left[\sum_0^{m+1} b_r z^{2r+1}\right] \quad \text{for} \quad z = e^{i\theta}.$$

Thus, by the uniqueness proved in (i), $g_n(z) = \sum_0^{m+1} b_r z^{2r+1} + ic$, where c is chosen to make $\Im[g_n(i)] = 0$. Observe that $\Re[g_n(i)] = 0$, by inspection.

(iii) Since z is a rational function of ζ, $g_n[z(\zeta)]$ is a rational function of ζ, that is, $f(\zeta)$ is rational.

(iv) $g_n(z)$ has a double root at $z = i$.

By (iii) $g_n(i) = 0$. We have to prove that $g_n'(i) = 0$.

At $z = i$, $\quad g_n'(z) = -\dfrac{\partial g_n}{\partial\theta}$; since $g_n'(i) = \sum_0^{m+1}(2r+1)b_r(-1)^r$, $g_n'(i)$ is real,

and
$$\frac{\partial g_n}{\partial\theta} = \frac{\partial\Re[g_n]}{\partial\theta} = \frac{\partial\mathfrak{F}_n}{\partial\theta}.$$

But
$$\frac{\partial\mathfrak{F}_n}{\partial\theta} = -A\cos\theta\sin\theta\{n\cos\theta[(1+\cos\theta)^{n-1} + (1-\cos\theta)^{n-1}] + 2[(1+\cos\theta)^n - (1-\cos\theta)^n]\},$$

so that for
$$\theta = \tfrac{1}{2}\pi, \quad \frac{\partial\mathfrak{F}_n}{\partial\theta} = 0.$$

(v) $f_n(\xi)$ is L_1.

Consider
$$\int_X^{X'} |f_n(\xi)|\,d\xi, \quad \text{which} = \int_{\Theta}^{\Theta'} |g_n(e^{i\theta})| \left|\frac{\partial\xi}{\partial\theta}\right| d\theta.$$

Since $|g_n(e^{i\theta})|$ is bounded, and the only singularity of $\partial\xi/\partial\theta$ is a pole $-2a/(z-i)^2$ at $\theta = \frac{1}{2}\pi$, where $g_n(z)$ has a double zero by (iv),

$$\int_{\Theta}^{\Theta'} |g_n(e^{i\theta})| \left|\frac{\partial\xi}{\partial\theta}\right| d\theta$$

is bounded. Hence $f_n(\xi)$ is L_1.

Note. (iii) and (iv) establish (20·6).

(vi) $\Re[f_n(\xi)]$ satisfies (19·3), i.e. $\Re[f_n(\xi)]$ is antisymmetrical and > 0 on I_i.

By (20·4) it is sufficient to consider $\mathfrak{F}_n(\theta)$.

$\mathfrak{F}_n(\theta)$ is antisymmetrical about $\theta = -\frac{1}{2}\pi$, since $\cos(-\frac{1}{2}\pi + \alpha) = -\cos(-\frac{1}{2}\pi - \alpha)$, so that $\cos^2\theta$ is symmetrical, $(1+\cos\theta)^n - (1-\cos\theta)^n$ is antisymmetrical about $\theta = -\frac{1}{2}\pi$. By (20·2) $\Re[f_n(\xi)]$ is antisymmetrical.

Also for
$$\tfrac{1}{2}\pi < \theta < \tfrac{3}{2}\pi, \quad (1-\cos\theta)^n > (1+\cos\theta)^n,$$

so that on,
$$\tfrac{1}{2}\pi < \pi - \delta < \theta < \pi + \delta < \tfrac{3}{2}\pi, \quad \mathfrak{F}_n(\theta) > 0.$$

Since the image of I_i is the arc $\pi - \delta \leqslant \arg z \leqslant \pi + \delta$, $\delta < \frac{1}{2}\pi$,

$$\Re[f_n(\xi)] > 0 \quad \text{on} \quad I_i.$$

(*c*) Lastly we show that if n is sufficiently large $\mathfrak{R}[f_n(\xi)]$ also satisfies (19·4), that is, it is small outside the intervals I_i, I_j.

Observe first (see Hardy & Rogosinski 1944, theorem 19): if $\epsilon > 0$, $\pi > \delta > 0$,

$$|\mathfrak{F}_n(\theta)| < \tfrac{1}{2}\epsilon\cos^2\theta \quad \text{on} \quad \delta \leqslant \theta \leqslant \pi - \delta$$

for sufficiently large n.

For
$$\frac{(1+\cos\theta)^n}{\int_{-\pi}^{\pi}(1+\cos\theta)^n\,d\theta} < \frac{2}{\delta}\left(\frac{\cos\frac{1}{2}\delta}{\cos\frac{1}{4}\delta}\right)^{2n} < \epsilon, \quad \text{for} \quad \delta \leqslant \theta \leqslant \pi,\ n > n(\delta),$$

and similarly
$$\frac{(1-\cos\theta)^n}{\int_{-\pi}^{\pi}(1+\cos\theta)^n\,d\theta} < \epsilon \quad \text{for} \quad 0 \leqslant \theta \leqslant \pi - \delta.$$

Since $|(1+\cos\theta)^n - (1-\cos\theta)^n| < \max(|1+\cos\theta|^n, |1-\cos\theta|^n)$, for $n > n(\delta)$,

and
$$\delta \leqslant \theta \leqslant \pi - \delta, \quad |\mathfrak{F}_n(\theta)| < \tfrac{1}{2}\epsilon\cos^2\theta.$$

Now
$$\int_{I_i}\mathfrak{R}[f_n(\xi)]\,d\xi = -\int_{I_j}\mathfrak{R}[f_n(\xi)]\,d\xi = -i\int_{-\delta}^{\delta}\mathfrak{F}_n(\theta)\frac{d\zeta}{dz}\mathrm{e}^{i\theta}\,d\theta.$$

Since
$$\frac{d\zeta}{dz} = -\frac{2a}{(z-i)^2}, \quad i\frac{d\zeta}{dz}\mathrm{e}^{i\theta} \to a \quad \text{as} \quad z \to 1,$$

$$-\int_{-\delta}^{\delta}\mathfrak{F}_n(\theta)\,d\theta \to 1 \quad \text{as} \quad n \to \infty, \quad \text{and therefore} \quad \int_{I_i}\mathfrak{R}[f_n(\xi)]\,d\xi \to a \quad \text{as} \quad n \to \infty,$$

$$\int_{J_{ij}}|\mathfrak{R}[f_n)]|\,d\xi = \int_{-\pi+\delta}^{-\delta} + \int_{\delta}^{\pi-\delta}|\mathfrak{F}_n(\theta)|\left|\frac{d\zeta}{dz}\right|d\theta < \epsilon\int_0^{2\pi}\frac{a\cos^2\theta\,d\theta}{|\cos\theta - i(1-\sin\theta)|}$$
$$= \tfrac{1}{2}a\epsilon\int_0^{2\pi}\frac{\cos^2\theta\,d\theta}{1-\sin\theta} = \tfrac{1}{2}a\epsilon\int_0^{2\pi}(1+\sin\theta)\,d\theta = a\pi\epsilon.$$

Summary of the construction of g_n, p, and ω

21. Suppose we have two intervals I_i, I_j and

$$k_i - c_i > k_j + c_j + \gamma, \quad \gamma > 0, \quad \text{or} \quad k_i + c_i + \gamma < k_j - c_j,$$

I_i to the left of I_j.

By lemma III (*a*) we determine a δ, depending on the length of the intervals and their mutual distance; $\sin\delta = (\rho_1 - \rho)/(\rho_1 + \rho)$, where 2ρ is the distance between their nearest, $2\rho_1$ between their farthest end-points.

By lemma III (*c*) we choose the least n so that

$$\frac{|(1+\cos\theta)^n - (1-\cos\theta)^n|}{\int_{-\pi}^{\pi}(1+\cos\theta)^n\,d\theta} < \frac{\gamma}{4\pi\max|k|}, \quad \text{in} \quad \delta \leqslant \theta \leqslant \pi - \delta.$$

Having fixed n we now choose the largest p_0, so that

$$\sum_{i=0}^{N}(|k_i| + c_i)\int_{I_i}|\mathfrak{I}\{g_n[z(\xi)]\}\sin p\xi|\,d\xi \leqslant \frac{\gamma}{8}\int_{I_i}\mathfrak{F}_n[\theta(\xi)]\,d\xi.$$

Then, for all $0 \leqslant p \leqslant p_0$, and any $k(\xi)$ consistent with $\{k_i, c_i\}$,

$$\left|\int_{-\infty}^{\infty} k(\xi)\, g_n[z(\xi)]\, e^{ip\xi}\, d\xi\right| > \frac{\gamma}{8}\int_{I_i} \mathfrak{F}_n[\theta(\xi)]\, d\xi = \omega > 0.$$

N.B. The problem of finding the best possible bound α_0 for α is still open, it is this: For any $\{k_i, c_i\}$ and $c > 0$, we are to find an α_0 so that

(1) for $\alpha > \alpha_0$, any ψ whose maximum modulus $< c$, is inconsistent with $\{k_i, c_i\}$;
(2) for $\alpha > \alpha_0$ we can construct a ψ, $|\psi| < c$, which is consistent with $\{k_i, c_i\}$.

Examples of theorem B

22. We return to the problem of gravity survey of para. 14 and use the notation introduced there. We ask: How near to Σ must the irregularities in the density ψ on Σ' be to account for the measured gravitational field k? That is, we want an upper bound for α so that $2G\alpha \int_{-\infty}^{\infty} \frac{\psi(x)\, dx}{(\xi - x)^2 + \alpha^2}$ be consistent with k.

To apply theorem B we require an upper bound for the *surface density* ψ; note that since the volume density is bounded this can be got if we can estimate the thickness of the layer Σ', where the irregularities in density are concentrated.

Apply theorem B to the integral equation (14·1). Observe that the kernel $L(\zeta) = \frac{2G\alpha}{\zeta^2 + \alpha^2}$ satisfies (17·1)—its only singularity in the upper half-plane is at $\zeta = i\alpha$, where its residue is $-iG$—so that, by para. 18, if $f(\zeta)$ is rational, regular in the upper half-plane, and L_1 on the real axis

$$\frac{e^{p\alpha}}{\int_{-\infty}^{\infty} |f(\xi + i\alpha)|\, d\xi} < \frac{2\pi G \max|\psi|}{\left|\int_{-\infty}^{\infty} k(\xi)\, f(\xi)\, e^{ip\xi}\, d\xi\right|}.$$

Since $$|f(\zeta)| = O|\zeta|^{-2}, \quad \int_{-\infty}^{\infty} |f(\xi + i\alpha)|\, d\xi = O(\alpha^{-1}),$$

so that the left-hand side of the inequality tends to infinity with α as $\alpha\, e^{p\alpha}$. The right-hand side is independent of α so that the inequality provides an upper bound for α.

23. It is easy to give bounds for the depth of the irreguliarities in density in terms of the *volume density* if we consider the field of masses distributed in the half-space $y > \alpha$.

Still considering a two-dimensional problem we measure y downwards from the plane $\Sigma\,(y = 0)$. Let the variations in volume density be $\psi(x, y)$, $\psi(x, y) = 0$ for $y < \alpha$, and suppose $k(\xi)$ is finite. Then

$$e^{p\alpha} < \frac{2\pi G \max|\psi|}{p} \frac{\max\limits_{y>0} \int_{-\infty}^{\infty} |f(x + iy)|\, dx}{\left|\int_{-\infty}^{\infty} k(\xi)\, f(\xi)\, e^{ip\xi}\, d\xi\right|}.$$

Proof. To avoid irrelevant complications due to the fact that, if $|\psi|$ is only required to be bounded at arbitrary depth, $k(\xi)$ need not exist at all, we work in terms of $k'(\xi)$ instead of $k(\xi)$. We express afterwards the result in terms of k:

$$k'(\xi) = -4G\int_{-\infty}^{\infty} dx \left(\int_{\alpha}^{\infty} \frac{(\xi-x)\, y\, \psi(x,y)\, dy}{[(\xi-x)^2+y^2]^2}\right).$$

Denote $\int_{\xi}^{\infty} f(u)\, \mathrm{e}^{ipu}\, du$ by $F(\xi, p)$. By contour integration we find that for $p \geqslant 0$, $F(-\infty) = 0$, and for $p>0$, $F(\xi) = O(|\xi|^{-2})$. By integration by parts

$$\int_{-\infty}^{\infty} k(\xi)\, f(\xi)\, \mathrm{e}^{ip\xi}\, d\xi = \int_{-\infty}^{\infty} k'(\xi)\, F(\xi, p)\, d\xi.$$

Substituting for k' from above we get

$$\int_{-\infty}^{\infty} k(\xi)\, f(\xi)\, \mathrm{e}^{ip\xi}\, d\xi = -4G\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{\alpha}^{\infty} \frac{(\xi-x)\, F(\xi)\, y\, \psi(x,y)}{[(\xi-x)^2+y^2]^2}\, dx\, d\xi\, dy,$$

where the order of integration is immaterial by absolute convergence of the integral.

Now

$$-4\int_{-\infty}^{\infty} \frac{(\xi-x)\, F(\xi)\, d\xi}{[(\xi-x)^2+y^2]^2} = 2\int_{-\infty}^{\infty} F(\xi)\, \frac{d}{d\xi}[(\xi-x)^2+y^2]^{-1}\, d\xi$$

$$= 2\int_{-\infty}^{\infty} \frac{f(\xi)\, \mathrm{e}^{ip\xi}\, d\xi}{(\xi-x)^2+y^2} = 2\pi \frac{f(x+iy)}{y}\, \mathrm{e}^{ipx}\, \mathrm{e}^{-py},$$

so that

$$\left|\int_{-\infty}^{\infty} k(\xi)\, f(\xi)\, \mathrm{e}^{ip\xi}\, d\xi\right| = 2\pi G \left|\int_{-\infty}^{\infty}\int_{\alpha}^{\infty} \psi(x,y)\, f(x+iy)\, \mathrm{e}^{ipx}\, \mathrm{e}^{-py}\, dx\, dy\right|$$

$$\leqslant \frac{2\pi G}{p}\, \mathrm{e}^{-p\alpha} \max_{y>\alpha} |\psi| \max_{y>\alpha} \int_{-\infty}^{\infty} |f(x+iy)|\, dx.$$

IV. Discussion of smoothing

In the present section we mention some further applications of the method of proof and of the results of theorem A.

24. The crucial step in the *proof* of theorem A is the elimination of the Fourier transform K of k by the expression (10·5) in para. 10 (i). This is also of use in various problems where the eigenfunctions are simply harmonic waves, e.g. in para. 29, and the answers are obtained in the form

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} KF\, \mathrm{e}^{i(p_1x_1+\dots+p_nx_n)}\, dp_1\dots dp_n, \tag{24·1}$$

where k is measured, and F is obtained theoretically. For if the Fourier transform f of F is L_1, and K is L_1 or L_2, by Titchmarsh (1937, theorems 35 and 65), (24·1) may be replaced by

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} k(\xi_1\dots\xi_n)\, f(x_1-\xi_1\dots x_n-\xi_n)\, d\xi_1\dots d\xi_n. \tag{24·2}$$

(24·2) is in practice superior to (24·1); it shows that the answer is determined continuously by k since f is L_1, while, if k can only be assumed to be L_1 or L_2, K, and therefore the integrand of (24·1), are not. In other words, K is not determined by measurements of k, but its 'smoothed version' (24·1) is. The virtue of the step from (24·1) to (24·2) is that it eliminates the difficulties due to the indeterminacy of K which is irrelevant to the result considered above.

Another use of the transformation is that (24·1) is derived for functions k which have a Fourier transform. (24·2) is at any rate *defined* for a wider class, e.g. for any bounded k, and it can often be shown to be the required result in the general case, cf. paras. 10 (ii) and 29 (*a*).

Restrictions on ψ

25. The *results* of theorem A can be used to determine ψ itself if rapidly oscillating ψ may be excluded, since then ψ is determined by $\overline{\psi}$. More precisely, we show:

If ψ is bounded ($|\psi| < c$), continuous at $(x_1 \ldots x_n)$, and, as in (12·1), if to any δ and ϵ, there is a μ satisfying (6·2), (8·2) and (8·3) so that

$$\int_{|x_1|+\ldots+|x_n|<\delta} \!\!\!\ldots\!\int |\mu(x_1 \ldots x_n)|\, dx_1 \ldots dx_n < C,$$

where C is an absolute constant, and

$$\int_{|x_1|+\ldots+|x_n|>\delta} \!\!\!\ldots\!\int |\mu(x_1 \ldots x_n)|\, dx_1 \ldots dx_n < \epsilon,$$

then $\psi(x_1 \ldots x_n)$ is determined continuously by k.

By the continuity of ψ, given $\eta > 0$, there is a δ so that

$$|\psi(x_1 \ldots x_n) - \psi(x_1' \ldots x_n)| < \tfrac{1}{2}\eta/C \quad \text{if} \quad |x_1' - x_1| + \ldots + |x_n' - x_n| < \delta.$$

Choose the μ corresponding to δ and $\tfrac{1}{4}\eta/c$. Then

$$\begin{aligned} &|\psi(x_1 \ldots x_n) - \overline{\psi}(x_1 \ldots x_n)| \\ &\quad = \left| \int_{-\infty}^{\infty} \!\ldots\! \int_{-\infty}^{\infty} [\psi(x_1 \ldots x_n) - \psi(u_1 \ldots u_n)]\, \mu(x_1 - u_1 \ldots x_n - u_n)\, du_1 \ldots du_n \right| \\ &\quad < \tfrac{1}{2}\eta C^{-1} \,.\, \int_{|u_1|+\ldots+|u_n|<\delta} \!\!\!\ldots\!\int \; + 2c \int_{|u_1|+\ldots+|u_n|>\delta} \!\!\!\ldots\!\int |\mu(u_1 \ldots u_n)|\, du_1 \ldots du_n < \eta. \end{aligned}$$

Since $\overline{\psi}$ is determined continuously by k, and η is arbitrary, this proves the result.

Note that the $\overline{\psi}_1$ constructed in para. 13 is continuous, so that for any $\{k_i, c_i\}$ there always exists a continuous ψ consistent with $\{k_i, c_i\}$.

Restriction on Ψ

26. Observe also that though generally the Fourier transform of measured functions comes into the statement of practical results only via an integral of the form (24·1), there are important applications of (1·1) where ψ may be so restricted that its asymptotic behaviour is expressed simply in terms of Ψ; e.g. take $n = 1$,

and suppose $\Psi\ (=\surd(2\pi)\,K/\Lambda)$ is known to be of the form $A(p)\,e^{i\chi(p)}$, where $A(p)$ is so smooth and small at infinity, $\chi(p)$ is known and so large when $p\to\infty$ that $\surd(2\pi)^{-1}\int_{-\infty}^{\infty}\Psi\,e^{ipx}\,dp$ may be evaluated by Kelvin's principle of stationary phase (Watson 1944, art. 8·2). Then, for large x,

$$\psi(x)\sim\left|\frac{d^2\chi}{dp^2}\right|^{-\frac{1}{2}}A(p_0)\,e^{i[\chi(p_0)+xp_0\pm\frac{1}{4}\pi]},$$

where $\chi'(p_0)+x=0$, and the last term in the exponent has the sign of $d^2\chi/dp^2{}_{p=p_0}$.

Now the value of the Fourier transform is itself of interest, and we therefore show, by para. 25, how it is to be derived from measurements of k.

27. By the Faltungstheorem suitable smoothed versions of the Fourier transform F of a function f are determined continuously by f, and F itself can be derived provided it is known to be smooth. In detail, for $n=1$:

Given a $c_0>0$, a $c>0$, a function $\delta(\epsilon)$ which $\to 0$ as $\epsilon\to 0$, and an $\eta>0$, we construct a function $g(x)$ of L_1, and an X, so that if

(*a*) $|f|<c_0$,
(*b*) its Fourier transform F exists, and $|F|<c$,
(*c*) f or F is L_1 or L_2,
(*d*) $|F(p_0+p)-F(p_0)|<\frac{1}{4}\eta C^{-1}$ for $|p|<\delta(\eta)$, then

$$\left|F(p_0)-\int_{-X}^{X}f(x)\,g(x)\,e^{ip_0x}\,dx\right|<\eta, \tag{27·1}$$

i.e. $F(p_0)$ is calculated from the values of $f(x)$ in the finite stretch $-X<x<X$.

For the proof we require that $g(x)$ is L_1 and L_2, and that its Fourier transform $G(p)$ satisfies

$$\int_{-\infty}^{\infty}G(p)\,dp=1,\quad \int_{-\delta(\eta)}^{\delta(\eta)}|G(p)|\,dp<C, \tag{27·2}$$

and

$$\int_{|p|>\delta(\eta)}|G(p)|\,dp<\tfrac{1}{8}\eta c^{-1}. \tag{27·3}$$

One such $g(x)=e^{-\alpha x^2}/\surd(2\pi)$ with sufficiently small α, since

$$G(p)=\tfrac{1}{2}\,e^{-\frac{1}{4}p^2/\alpha}/\surd(2\alpha\pi),\quad \int_{-\delta(\eta)}^{\delta(\eta)}|G(p)|\,dp<1,$$

and for given $\delta(\eta)$, $\int_{|p|>\delta(\eta)}|G(p)|\,dp$ can be made arbitrarily small by choosing α sufficiently small.

Observe that the smoothed version $\int_{-\infty}^{\infty}F(p)\,G(p_0-p)\,dp$ of $F(p)$

$$=\int_{-\infty}^{\infty}f(x)\,g(x)\,e^{ip_0x}\,dx, \tag{27·4}$$

by Titchmarsh (1937, theorem 40), if f or F is L_1 since g and G are L_1, and by theorem 64 if f is L_2 since g is L_2. It is determined continuously by (27·4), since g is L_1 and f is bounded.

Substituting $G(p)$ for μ in para. 25 we find that, by (27·2) and (27·3), the smoothed version of $F(p)$ differs from $F(p_0)$ by less than $\frac{1}{2}\eta$, so that by (27·4)

$$\left| F(p_0) - \int_{-\infty}^{\infty} f(x)\, g(x)\, \mathrm{e}^{ip_0 x}\, dx \right| < \tfrac{1}{2}\eta. \tag{27·5}$$

Also $$\left| \int_{-\infty}^{\infty} f(x)\, g(x)\, \mathrm{e}^{ip_0 x}\, dx - \int_{-X}^{X} f(x)\, g(x)\, \mathrm{e}^{ip_0 x}\, dx \right| < c_0 \int_{|x|>X} |\, g(x)\,|\, dx,$$

which can be made less than $\frac{1}{2}\eta$ by choosing X sufficiently large, since g is L_1. Thus, by (27·5), the theorem follows.

Note that if $F(p) = A(p)\, \mathrm{e}^{i\chi(p)}$ as in the preceding paragraph, we may take instead of $G(p)$ the function $G_1(p) = G(p)\, \mathrm{e}^{-i\chi(p_0-p)}$ if $\int_{-\infty}^{\infty} G(p)\, \mathrm{e}^{-i\chi(p_0-p)}\, dp \neq 0$, and if its Fourier transform is L_1, since $G_1(p)$ satisfies (27·2) and (27·3).

Example

28. To illustrate the remark of para. 24, and to compare the results of paras. 25 and 26 with what can be obtained by theorem A, we turn to a problem in the theory of progressive surface waves in deep water.

By a use of the Faltungstheorem we derive a relation of type (1·1) by which the pressure at a point Q of the water is calculated from the pressure at P when P is above Q. As in para. 3, from measurements at Q we generally only get a smoothed version of the pressure at P.

The much more simple and complete solution of para. 26 applies when the motion is generated by pressures over a region whose diameter is small compared with its distance from P. With this extra condition the motion at P is continuously determined by measurements at Q. The results are particularly simple if both P and Q are far from the area of disturbance, and P is beyond Q.

Proof. 29. (*a*) *Relation between the pressures at P and Q.* Recall that in two-dimensional progressive wave motion in the (x, y) plane, where $y = 0$ denotes the mean free surface, and y is measured vertically downwards, the pressure $\pi(x, y;\, t)$ is given by

$$\frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} A(p)\, \mathrm{e}^{i(\operatorname{sgn} p)\, p^2 x/g}\, \mathrm{e}^{-p^2 y/g}\, \mathrm{e}^{ipt}\, dp \quad (y > 0), \tag{29·1}$$

by Lamb (1932, art. 229); (29·1) can be established with $A(p)/(1 + |\, p\,|)^{\frac{1}{2}}$ belonging to L_2 if the energy of the motion and the vertical velocity are bounded; also π is L_2 with respect to t when $y > 0$.

Denoting the Fourier transform of π with respect to t by Π we get from (29·1)

$$\Pi(\xi, \eta;\, p) = \Pi(x, y;\, p)\, \mathrm{e}^{i(\operatorname{sgn} p)\, p^2(\xi - x)/g}\, \mathrm{e}^{-p^2(\eta - y)/g}. \tag{29·2}$$

By the Faltungstheorem, applied to (29·2), for $\eta > y > 0$,

$$\pi(\xi, \eta;\, \tau) = \int_{-\infty}^{\infty} \pi(x, y;\, t)\, L_0(\xi - x, \eta - y;\, \tau - t)\, dt, \tag{29·3}$$

where $$L_0(u, v;\, w) = \frac{1}{\pi} \int_0^{\infty} \mathrm{e}^{-p^2 v/g} \cos\, (p^2 u/g + wp)\, dp. \tag{29·4}$$

Note that L_0 is L_1 with respect to w, by repeated integration by parts of (29·4), as required.

(*b*) Smoothed versions of $\pi(x, y; t)$ are obtained from $\pi(\xi, \eta; \tau)$ by a straightforward use of theorem A.

Suitable smoothing functions have the Fourier transforms $M(p) = \sqrt{(2\pi)^{-1}}\, \mathrm{e}^{-p^4/\beta}$, $\beta > 0$; they satisfy the conditions of para. 25 for the kernel (29·4) when β is sufficiently large. Therefore the pressure at (x, y) is determined continuously by the pressure at (ξ, η) if $\pi(x, y; t)$ is continuous with respect to t.

(*c*) Note that though (29·1) is derived for π of L_2 only, it can be shown that (29·3), holds for any bounded $\pi(x, y; t)$.

30. We restrict π further in order to apply paras. 25 and 26.

(i) A bound for the modulus of continuity of $\pi(x, y; t)$ with respect to t is obtained if it may be assumed that

$$|\pi(x, 0; 0)| < M. \tag{30·1}$$

(ii) If also $\pi(x, 0; 0) = 0$ for

$$|x| > A, \tag{30·2}$$

the pressure integral (29·1) is given asymptotically, i.e. for A/x small, by

$$-\sqrt{(2g/x)}\, |A(p_0)| \cos\left(\tfrac{1}{4}gt^2/x + \alpha\right) \mathrm{e}^{-\frac{1}{4}gt^2/x}, \quad p_0 = -\tfrac{1}{2}gt/x, \tag{30·3}$$

where $$\alpha = \arg A(p_0) - \tfrac{1}{4}\pi.$$

(iii) If gt^2/x is not too large, for fixed x, $A(p_0)$ varies little during a period $T = 8\pi x/(gt)$, so that (30·3) is nearly a sine wave over several multiples of T.

(*a*) Under the condition (30·1) it is convenient to replace (29·3) by

$$\pi(\xi, \eta; \tau) = \int_{-\infty}^{\infty} \pi(x, y; t)\, L_0^*(\xi - x, \eta - y; \tau - t)\, dx, \tag{30·4}$$

where $$L_0^*(u, v; w) = \frac{2}{\pi} \int_0^{\infty} p\, \mathrm{e}^{-p^2 v/g} \cos\left(p^2 u/g + wp\right) dp. \tag{30·5}$$

(To derive (30·4) from (29·1) put $p^2 \operatorname{sgn} p = q$ in (29·1) so that the resulting integral is a Fourier integral with respect to x, and then apply the Faltungstheorem as in 29 (*a*).)

(*b*) Under (30·1) $\pi(x, y; t)$ is plainly continuous with respect to t since

$$\int_{-\infty}^{\infty} |L_0^*(u, v; w)|\, du$$

is uniformly continuous with respect to w; write

$$\frac{\partial L_0^*}{\partial w} = \frac{gw}{2u}\left[\int_0^{\infty} p\, \mathrm{e}^{-p^2 v/g} \sin pw\, d \sin\left(p^2 u/g\right) - \int_0^{\infty} p\, \mathrm{e}^{-p^2 v/g} \cos pw\, d \cos\left(p^2 u/g\right)\right],$$

and integrate by parts twice.

Thus by 29 (*b*) $\pi(x, y; t)$ is continuously determined by $\pi(\xi, \eta; \tau)$.

(*c*) When

$$\pi(x, y; t) = \int_{-A}^{A} \pi(u, 0; 0)\, L_0^*(x - u, y; t)\, du,$$

we derive the expression (30·3) from an asymptotic expression for L_0^* obtained by Kelvin's principle; for details see Lamb (1932, p. 395). The expression is valid uniformly in $x-A<u<x+A$ and has an error $O(A/x)$.

(*d*) To establish 30 (iii) we have to show that $A(p_0)$ varies little during a period $8\pi x/gt$.

Note that by (29·1)

$$gA(p) = p\sqrt{(2\pi)^{-1}}\int_{-A}^{A}\pi(u,0;\,0)\,e^{-ip^2u/g}\,du.$$

For fixed x,

$$\begin{aligned}\frac{\partial A}{\partial t} &\leqslant \sqrt{(2\pi)^{-1}}\,[MA+p^2MA^2/g]/x, \quad \text{since} \quad \left|\frac{\partial p}{\partial t}\right| = \tfrac{1}{2}g/x,\\ &= \sqrt{(2\pi)^{-1}}\,M(A/x)\,[1+\tfrac{1}{4}(gt^2/x)\,(A/x)].\end{aligned}$$

Thus if A/x and $(gt^2/x)(A/x)$ are both small, the motion near, and beyond, P is approximately simple harmonic.

This completes the proof of the statements of para. 28.

REFERENCES

Bullard, E. C. & Cooper, R. I. B. 1948 *Proc. Roy. Soc.* A, **194**, 332–347.
Doss, R. 1948 *Proc. Lond. Math. Soc.* (2), **50**, 169–195.
Hardy, G. H. & Rogosinsky, W. W. 1944 *Fourier series*. Cambridge University Press.
Lamb, H. 1932 *Hydrodynamics*, 6th ed. Cambridge University Press.
Titchmarsh, E. C. 1937 *Introduction to the theory of Fourier integrals*. Oxford.
Watson, G. N. 1944 *A treatise on the theory of Bessel functions*, 2nd ed. Cambridge University Press.

The catalytic decomposition of simple heterocyclic compounds

I. Preparation and properties of molybdenum sulphide and nickel sulphide catalysts

BY E. H. M. BADGER, R. H. GRIFFITH AND W. B. S. NEWLING
Fulham Laboratory, The Gas Light and Coke Co.

(*Communicated by E. J. Bowen, F.R.S.—Received* 9 *July* 1948—
Revised 3 *December* 1948—*Read* 17 *March* 1949)

The preparation and properties of molybdenum disulphide and nickel subsulphide catalysts, used for the decomposition of simple heterocyclic compounds, are described. Equilibria between nickel sulphide, nickel subsulphide and metallic nickel have also been examined by application of magnetic measurements. The adsorption of hydrogen and of the heterocyclic compounds has been determined on the various catalysts over a wide temperature range.

INTRODUCTION

Interest in the catalytic disruption of five-membered heterocyclic compounds has been largely confined to the behaviour of thiophen; this is a consequence of the frequent occurrence of this substance in gaseous and liquid fuels, and of the desirability for its removal from such mixtures. While some success has attended efforts to decompose thiophen by catalytic hydrogenation under pressure, attempts to apply similar catalysts at normal pressures to the removal of thiophen from coal gas have not been successful. In pure hydrogen, however, the breakdown of thiophen at normal pressure occurs with sulphide catalysts containing molybdenum or nickel.

The present investigation was undertaken to study the mechanism of the breakdown of thiophen, furane and pyrrole by catalytic hydrogenation at normal pressure; the oxygen and nitrogen compounds were included in the hope that differences in the behaviour of the three would give valuable information.

This paper describes the preparation and physical properties of molybdenum sulphide and nickel sulphide catalysts. It also includes information on the adsorption of hydrogen and of heterocyclic compounds, and measurements of chemical equilibria in the nickel sulphide-nickel hydrogen system. The following paper is concerned with the reaction kinetics of the catalytic decomposition of the heterocyclic compounds, and proposes two separate mechanisms by which molybdenum sulphide and nickel subsulphide respectively function.

EXPERIMENTAL

(1) *The preparation and properties of nickel sulphide catalysts*

Preparation of the normal sulphide

Nickel sulphide is commonly prepared by precipitation, but the product is impure (Middleton & Ward 1935). For catalytic work, a satisfactory material is obtained by treating finely divided nickel with H_2S or CS_2 vapour with close temperature control to prevent sintering of the nickel. Less temperature control is needed for the direct reduction of nickel sulphate by gases rich in hydrogen, a reaction discovered by Arfvedson (1825) who noted the evolution first of SO_2, then of H_2S and described the product as Ni_2S. Recognition of the sulphate reduction as a method for the preparation of active Ni_3S_2 catalysts is relatively recent (Griffith & Plant 1939).

It has now been found that the primary reaction $NiSO_4 + 4H_2 = NiS + 4H_2O$ starts at a minimum temperature of about 350° C whether in coal gas or pure hydrogen. It is also found that the secondary reduction of NiS to Ni_3S_2 is a slower process, and that the H_2S produced reacts rapidly with unchanged sulphate liberating an equal amount of SO_2 and yielding more NiS, as in the equation

$$3NiSO_4 + 4H_2S = 3NiS + 4SO_2 + 4H_2O.$$

Nickel sulphide, free from subsulphide, can be prepared by this reaction using gases rich in H_2S; the action of the H_2S starts at about 165° C, and the range 200 to 250° C is suitable for complete reduction of the sulphate.

The most convenient preparation of NiS is by the treatment of the hydroxide with H_2S. The reaction is instantaneous at room temperature, so that the rate of sulphiding is limited only by the access of H_2S to the unconverted hydroxide. The conditions should therefore permit rapid evaporation of the water produced. In a static system, pure H_2S may be used undiluted at low pressure. At atmospheric temperature and pressure, the H_2S should be much diluted by a dry gas and used in a streaming system. At 100° C little dilution of the H_2S stream is necessary. The hydroxide is readily formed into small tablets of uniform shape and size in a suitable reciprocating machine, or a hard press-cake may be broken and graded. Either granular form may be sulphided with little shrinkage or loss of shape to give a mechanically strong product of about 50 % porosity by volume; the extent of the large internal surface decreases considerably with increased temperature of sulphiding. Basic nickel carbonate may be similarly treated, but the resulting sulphide is excessively porous and reactive, and mechanically rather weak.

For the experimental work on adsorption and rates of hydrogenation of heterocyclic compounds, the nickel sulphide catalysts used were made from a stock of $Ni(OH)_2$, which was quite free from soluble matter, but contained traces of Co and Fe. This stock was formed into $\frac{1}{4}$ in. diam. tablets in a pharmaceutical machine at about 300 atm. pressure, and the tablets broken into pieces of 1·5 to 2 mm. (screened 8–12 mesh B.S.S.). The graded material was sulphided in the adsorption apparatus or reaction tube at 85 to 90° C with hydrogen sulphide made by the action of hydrochloric acid on ferrous sulphide; the hydrogen sulphide was freed from arsenic, oxygen, acid gases and water vapour by the reagent train—solid $CaCl_2$, solid I_2, $CrCl_2$ solution, KSH solution and solid $Mg(ClO_4)_2$. A smaller stock of pure nickel hydroxide, specially free from iron, was the source of materials used in the magnetic measurements, in measurements of the reduction equilibria and for X-ray analysis; this was sulphided in each case at 100° C in pure H_2S made in a Meade generator

(Meade 1938) from pure alkali sulphide solution and sulphuric acid; the hydrogen sulphide was washed with $CrCl_2$ and KSH solutions and dried with solid $CaCl_2$ and $Mg(ClO_4)_2$.

In all methods of NiS preparation using H_2S as a reactant, it is essential that oxygen should be excluded. Oxygen reacts catalytically with H_2S on the outer skin of NiS even at room temperature, and the resulting sulphur can prevent completely the access of H_2S to unconverted material inside the granule, while at 100° C part of the sulphur formed becomes chemisorbed on the NiS surface and cannot be removed by simple physical methods. Any preparation of NiS which is required free from elementary sulphur should be evacuated or purged with CO_2 to remove all physically adsorbed H_2S before exposure to air. Only the cold sulphide should be exposed, and then with caution; on sudden exposure, the exothermic chemisorption of oxygen on the NiS can raise the temperature enough to cause ignition, the products being SO_2, NiO and $NiSO_4$.

The properties of nickel sulphide

The chemisorption of oxygen is the outstanding property of NiS. Even at −78° C the sorption is almost instantaneous and rapidly reaches a saturation value. Apart from a small proportion physically adsorbed, the amount of oxygen taken up is independent of temperature and so firmly held that it cannot be removed by evacuation, except as SO_2 at temperatures high enough for the oxidation of NiS (e.g. 140 to 200° C). One sample of NiS made from basic carbonate at about 20° C took up 60 ml. O_2/g. or 8·5 % by weight. A similar preparation made from $Ni(OH)_2$ adsorbed 39 ml. O_2/g. or 5·6 % by weight, but when the hydroxide was sulphided at 90° C the irreversible adsorption of oxygen was 11·2 ml./g. or 1·6 % by weight. The latter material had a specific surface of 36 m.^{2}/g., equivalent to 8·2 ml. N_2/g. of NiS for a monomolecular layer.

X-ray examination of nickel sulphide, prepared from iron-free hydroxide at 100° C and transferred without exposure to air into a thin-walled capillary tube of Lindemann glass (Badger 1947), showed the structure of the hexagonal form, β-NiS (Levi & Baroni 1935). Nickel sulphide, similarly prepared, but saturated with chemisorbed oxygen or sulphur, had the same structure but with slight expansion of the lattice parameters. The expansion of the lattice may be considered together with the relative adsorptions of atomic oxygen and of molecular nitrogen for a monolayer. The larger amount of oxygen could be due to particularly close and orderly packing in a monolayer, or there could be penetration of oxygen into the lattice. The considerable lowering of the free-surface energy of nickel sulphide, occurring when oxygen or sulphur is chemisorbed, would be sufficient to account for the lattice expansion of very small crystals. Although the X-ray diffraction lines from both the oxygen-free and the oxygen-laden nickel sulphide indicated a grain size of the order of 10^{-4} cm., which is relatively small, there was no evidence in either case of the marked range of lattice spacing value which would be expected to result from a superficial effect. It is therefore probable that the increase of the crystal parameters, occurring when oxygen is chemisorbed, is due to penetration of some oxygen into the bulk of the material.

The reduction of NiS *to* Ni_3S_2

In the absence of oxygen, the reduction of NiS by hydrogen begins below 100° C. When the sulphide has been in contact with oxygen, the reduction cannot be started below 130° C, at which temperature adsorbed oxygen is removed as water vapour by reaction with the hydrogen. Complete reduction to the subsulphide is readily effected at 250° C; above 360° C metallic nickel begins to appear.

The specimens of Ni_3S_2 prepared at 360° C had the structure described by Westgren (1938). All the X-ray examinations of the nickel sulphide and nickel subsulphide catalysts were carried out by the National Physical Laboratory.

The magnetic properties of nickel sulphide catalysts

The magnetic susceptibility of NiS is very low, values of $2\cdot5$ to $3\cdot5 \times 10^{-6}$ being recorded for the mass susceptibility (Klemm & Schüth 1933). That of the subsulphide was not known. Measurements were made in the hope that the subsulphide was sufficiently magnetic to provide a means of studying the reduction equilibrium:

$$3NiS + H_2 \rightleftharpoons Ni_3S_2 + H_2S.$$

NiS was prepared from the hydroxide at 100° C, and the mass susceptibility determined with the Gouy balance, giving:

$\chi = 2\cdot1 \times 10^{-6}$ e.m.u./g. for pure NiS prepared in the absence of air;
$\chi = 6\cdot0 \times 10^{-6}$ e.m.u./g. for NiS saturated with adsorbed oxygen.

The reduction of the same NiS stock was accompanied by more loss in weight than would be expected for subsulphide formation, and the product exhibited ferromagnetism which was destroyed on treatment with CO at 110 to 130° C. Metallic nickel was clearly present in the reduced catalyst. The marked increase of apparent susceptibility, which indicates the presence of those nickel aggregates, of 1000 or more atoms, that display ferromagnetism (Selwood 1946), does not occur below 350° C and requires higher temperatures with some samples; some finely dispersed nickel is probably formed at lower temperatures, while above 400° C part of the nickel formed will not react with CO.

2·5 g. of dry NiS saturated with adsorbed oxygen were subjected successively to the treatments shown in table 1, being at each stage cooled and exposed cautiously to air before measurement of susceptibility; previous experiment in a totally enclosed system had shown that this exposure caused no change of susceptibility.

TABLE 1

successive treatments (gas rate 3·5 l./hr.)			susceptibility $\times 10^6$ at 1660 oersted after treatment
gas	duration (hr.)	temperature (° C)	
H_2	$5\frac{1}{3}$	300	5·4
H_2	$5\frac{1}{3}$	300	7·1
H_2	3	370	171
H_2	$6\frac{1}{2}$	370	341
CO	$6\frac{3}{4}$	115	9·1
CO	6	115	7·0

The CO treatment was preceded by reduction of the material in H_2 at 200° C to remove oxygen, and subsequent cooling in the H_2 to 115° C.

Exhaustive treatment with CO gave Ni_3S_2 free from metallic nickel and of mass susceptibility $\chi = 4{\cdot}3 \times 10^{-6}$ e.m.u./g.

Equilibria in the reduction of nickel sulphides

No previous work has been published concerning equilibria in the reduction of Ni_3S_2 to the metal. Some figures are available for the H_2S formed when H_2 is passed over NiS (Jellinek & Zakowski 1925), but the presence of Ni_3S_2 in the reduction product was not recognized, and the conditions were not those of true equilibrium.

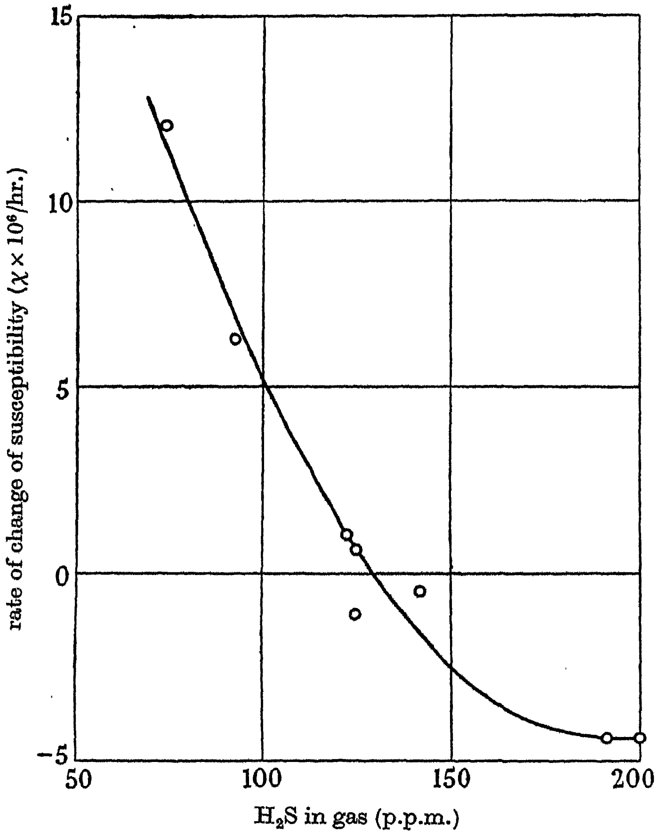


FIGURE 1. The rate of change of the magnetic susceptibility of a Ni_3S_2-Ni mixture in H_2-H_2S mixtures at 450° C.

The detection of an increase or decrease in the amount of nickel present by means of magnetic susceptibility measurements made it possible to determine whether a stream of hydrogen containing a constant known amount of H_2S was sulphiding or reducing a mixture of Ni_3S_2 and Ni. At a given temperature a gas stream containing the equilibrium concentration of H_2 and H_2S would naturally produce no change in the solid phase. It was therefore sufficient to vary the composition of the gas until no change in the magnetic susceptibility occurred.

To prepare the gas mixtures, hydrogen from a cylinder was passed at a constant rate through $CrCl_2$ solution and then over CuO heated to 600° C. Part of the oxygen-free hydrogen was diverted, also at a constant rate, through a bubbler containing a buffered sodium sulphide solution at room temperature; this added 200 p.p.m. (parts per million by volume) of H_2S to the stream which was then mixed with the rest of the pure hydrogen; the exact concentration of H_2S required was obtained by varying the fraction of hydrogen passing through the buffer solution. This contained 120 g. of $Na_2S,9H_2O$, 71 g. of Na_2HPO_4 and 5 ml. of phosphoric acid (sp.gr. 1·75) per litre of aqueous solution and had a pH of 10·6. The concentration of H_2S was measured by passing the mixed gas stream through cadmium acetate solution and determining the cadmium sulphide by means of standard iodine solution.

Figure 1 shows the rates of change of magnetic susceptibility at 450° C of a mixture of Ni_3S_2 and Ni which had been prepared by the reduction of NiS in hydrogen, plotted against the concentration of H_2S in hydrogen. The concentration of H_2S for no change in the susceptibility of the solid was 128 p.p.m., which is therefore the equilibrium value corresponding to 1 atm. of hydrogen. Similar measurements at 350° C showed that the equilibrium concentration of H_2S was 50 p.p.m.

An approximate value for the equilibrium pressure of H_2S for the reaction $3NiS + H_2 \rightleftharpoons Ni_3S_2 + H_2S$ was obtained by passing a very slow stream of H_2 over a large excess of NiS at 360° C; this gave a steady outlet concentration of 44·5 % H_2S by volume. It is therefore evident that only Ni_3S_2, or mixtures of this sulphide with metallic nickel, are stable under the conditions of catalytic decomposition of the heterocyclic compounds by gases rich in hydrogen.

(2) *Adsorption on* Ni_3S_2 *catalysts*

Adsorption measurements were carried out on all the catalysts at low temperatures to determine their total surface by the modified Brunauer, Emmett & Teller (1938) method, and at temperatures up to 400° C in order to study activated adsorption. In the latter series, determinations with thiophen, furane, pyrrole and their tetrahydro derivatives were carried out by the method described by Griffith, Hill & Plant (1937) using a steam-jacketed McLeod gauge for pressure measurements. Tetrahydrothiophen was prepared by the action of NaSH on 1.4-dichlorbutane; the other heterocyclic compounds were commercial products purified by fractional distillation.

The catalysts were prepared in a small glass bulb, through which a stream of gas could be passed as required, the outlet tube being sealed before evacuation of the catalyst was started. In this way exposure to air could not occur.

Ni_3S_2 *prepared at* 300° C

In a previous publication (Griffith & Hill 1938) the adsorption of hydrogen on nickel subsulphide was reported to be small at temperatures up to 350° C. New measurements on a specimen prepared by reduction of NiS at 300° C and evacuated for 3 days at 300° C, both before and after measurement of surface area, gave an area of 10·3 m.²/g. and hydrogen adsorption about ten times the previous value, indicating that in the earlier work hydrogen adsorbed during the preparation of the subsulphide had not been properly removed.

On similar specimens of Ni_3S_2, the adsorptions of thiophen, furanare, tethydrothiophen and tetrahydrofurane were measured. The results are shown in table 2. All the determinations were carried out with rising temperatures from 0 to 300° C. On repeating the observations with falling temperatures, all the adsorptions showed normal isobars with increasing values at the lower temperatures.

TABLE 2. ADSORPTION ON Ni_3S_2 PREPARED AT 300° C (ML./G.)

(Specific surface = 10·3 m.^{2}/g.)

temperature (° C)	hydrogen at 670 mm.	thiophen at 25 mm.	furane at 34 mm.	tetrahydrothiophen at 11 mm.	tetrahydrofurane at 24 mm.
0	0	—	1·7	0·86	1·95
66	—	—	0·27	0·55	1·16
100	0	0·08	0·20	0·53	0·92
150	—	0·045	0·10	0·49	0·64
200	—	0·15	0·09	0·20	0·29
250	4·6	0·105	—	−0·01	0·24
300	9·5	—	0·09	—	0·08

These figures show that although activated adsorption of the heterocyclic compounds does occur, it is extremely limited in extent. Comparable results for CS_2 (Griffith & Hill 1938) show about 100 times as much activated adsorption.

Ni_3S_2 *prepared at* 400° C

NiS reduced at 400° C for 24 hr. was strongly magnetic and had a surface area of 7·42 m.^{2}/g. The adsorption of both thiophen and furane was measured on this material.

TABLE 3. ADSORPTION OF THIOPHEN ON Ni_3S_2 PREPARED AT 400° C

(Specific surface = 7·42 m.^{2}/g.)

temperature (° C)	thiophen (ml./g.) at 25 mm.	temperature (° C)	thiophen (ml./g.) at 25 mm.
0	2·55	200	0·65
66	0·85	250	0·64
100	0·70	300	0·61
150	0·66	400	0·54

On repetition of the measurements after thorough evacuation of the catalyst, somewhat lower amounts were taken up, and it is probable that gradual reconversion of nickel to its sulphide was taking place.

These results suggest that thiophen reacts easily with part of the catalyst, so that the observations no longer represent true adsorption and are materially higher than those given by the catalyst prepared at 300° C.

With furane, there are indications of extensive decomposition, especially at 400° C, and this effect becomes evident at 100° C in a second adsorption after thorough evacuation. This suggests that an appreciable change has occurred in the catalyst and makes quantitative measurements of little value, but emphasizes the difference between the catalysts prepared at 300 and 400° C respectively.

Reaction of thiophen and furane with metallic nickel

A nickel subsulphide catalyst known to contain metallic nickel, with a magnetic mass susceptibility of 176×10^{-6}, was treated at 400° C with a stream of nitrogen containing 728 p.p.m. of thiophen for $3\frac{1}{2}$ hr. The susceptibility fell to 97×10^{-6}. A similar treatment with 760 p.p.m. of furane only reduced the susceptibility to 117×10^{-6}. This indicates that both heterocyclic compounds react with nickel, and that thiophen does so more easily.

(3) *Molybdenum sulphide catalysts*

Two methods are available for preparing active molybdenum disulphide catalysts: (1) by the action of H_2S on molybdic acid; (2) by the reduction of MoS_3 in hydrogen. The second method is recognized as giving the disulphide MoS_2, but the products obtained by the first procedure are far less well defined.

Galle & Michelitsch (1935) stated that a definite oxysulphide MoOS was produced by the action of H_2S on MoO_3. It therefore appeared to be important to establish whether such a compound actually occurred in any of the catalysts prepared for use in the hydrogenation of heterocyclic organic substances.

Specimens were accordingly made by the method of Galle & Michelitsch, and by modifications of this process, and examined by X-ray analysis. The results showed clearly that the products consisted of crystals of MoO_2 and MoS_2, together with amorphous material, and no evidence could be found for the existence of MoOS.

As far as the catalysts prepared from molybdic acid are concerned, only the structures of MoO_2 and MoS_2 have therefore to be considered.

(*a*) *Preparation of mixed oxide-sulphide catalysts*

Molybdic acid (approximate composition MoO_3, 0·96 H_2O) in powder form was compressed into large tablets at 2000 atm. pressure, and these were broken into pieces about $\frac{1}{4}$ in.; 800 g. of the granular material were heated for 150 hr. at 100° C in a stream of H_2S and then reduced in hydrogen at 400° C for 150 hr. to give a metallic grey product which was broken and graded. Granules of 1·5 to 2 mm. (8–12 B.S.S.) were used for both the reaction rate measurements and the adsorption experiments. Analysis of this material showed it to contain 60 % Mo and 28 % S by weight; it is therefore a mixture of oxide and sulphide.

(*b*) *Preparation of the disulphide by reduction of the trisulphide*

Ammonium molybdate in aqueous ammonia was treated with H_2S to precipitate ammonium thiomolybdate, which was filtered off, redissolved in water and decomposed by dilute aqueous sulphuric acid, as described by Vanino (1925). The precipitated MoS_3 was dried at 40° C, compressed into tablets to 2000 atm., broken, graded and reduced in hydrogen at 400° C. Again granules of 1·5 to 2 mm. were used.

X-ray analysis of a separately reduced specimen showed some of the lines expected from the structure of MoS_2. A considerable amount of amorphous material was present. The catalyst was considered to have a structure resembling that described by Mering & Levialdi (1941) for specimens obtained by the thermal decomposition of ammonium thiomolybdate. These had the basic lattice of molybdenite, but were defect structures in which a deficiency of molybdenum was associated with the progressive disappearance of reflexions from planes inclined to the threefold axis.

The X-ray examinations of the molybdenum sulphide catalysts, like those of the nickel sulphide series, were carried out at the National Physical Laboratory, and the results of these investigations are published by permission of the Director of the Laboratory.

(4) *Adsorption on molybdenum sulphides*

Two catalysts were prepared as described in § 3 and will be called MoO_2-MoS_2 and MoS_2. The surfaces were cleaned by heating in hydrogen for 6 hr. at 400° C, followed by evacuation at 400° C for 48 hr.

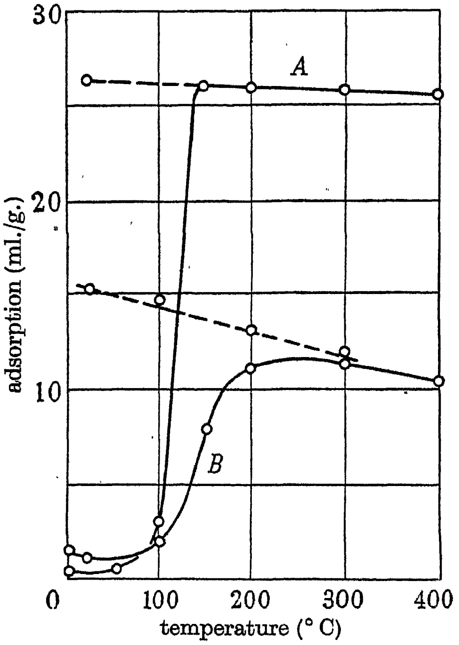


FIGURE 2. Isobars of hydrogen adsorption on molybdenum sulphide catalysts. *A*, MoS_2; *B*, MoO_2-MoS_2.

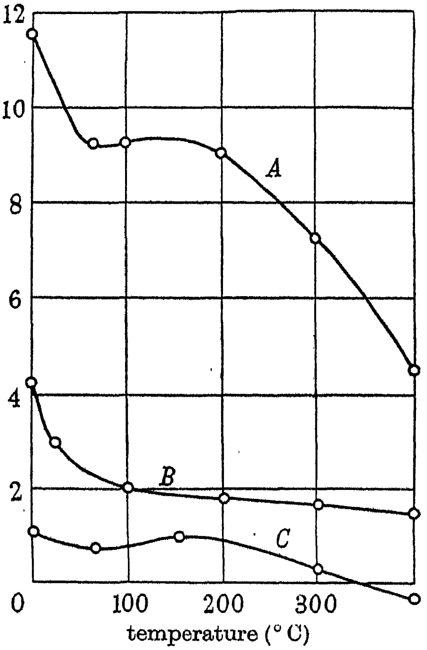


FIGURE 3. Isobars for heterocyclic compounds on molybdenum sulphide catalysts. *A*, thiophen on MoO_2-MoS_2; *B*, thiophen on MoS_2; *C*, tetrahydrothiophen on MoS_2.

The surface areas were 16·9 m.²/g. for MoO_2-MoS_2 and 16·6 m.²/g. for MoS_2.

Figure 2 shows isobars for hydrogen adsorption on the two catalysts.

Figure 3 shows isobars for thiophen adsorption on the two catalysts and for tetrahydrothiophen on MoS_2. Table 4 contains values for the adsorption of furane and tetrahydrofurane on MoS_2. For the heterocyclic compounds the curves were obtained by increasing the temperature after each reading and waiting ½ hr. before taking the next, so that an isobar was completed in 1 day.

The adsorbed tetrahydro derivatives decomposed at the higher temperatures and at 400° C extensive decomposition continued, giving an increase in pressure. Furane was appreciably more stable and thiophen was hardly decomposed at all.

After evacuation at 400° C for 2 or 3 days the curves could be repeated, giving slightly less adsorption.

Hydrogen adsorption took place more slowly, and the readings were obtained by observing the pressure over a number of hours until no further change occurred. The curves are beautiful examples of a reversible activated adsorption unobscured by physically adsorbed material.

TABLE 4. ADSORPTION ON MoS_2 (ML./G.)

(Surface area = 16·6 m.^{2}/g.)

temperature (° C)	0	65	100	150	200	300	400
furane at 34 mm.	2·6	1·2	1·0	1·0	1·0	1·05	0·8
tetrohydrofurane at 24 mm.	2·0	0·8	0·7	0·6	0·6	0·1	−0·2

In the work here described we have received much assistance from Miss P. R. Lindars and from Mr P. R. Chapman.

REFERENCES

Arfvedson, J. A. 1825 *Quart. J. Sci.* **18**, 392.
Badger, E. H. M. 1947 *J. Sci. Instrum.* **24**, 109.
Brunauer, S., Emmett, P. H. & Teller, E. 1938 *J. Amer. Chem. Soc.* **60**, 309.
Emmett, P. H. 1945 *Industr. Engng Chem.* **37**, 639.
Galle, E. & Michelitsch, W. 1935 *Montan. Rdsch.* **27**, no. 5, 1.
Griffith, R. H. & Hill, S. G. 1938 *J. Chem. Soc.* p. 717.
Griffith, R. H., Hill, S. G. & Plant, J. H. G. 1937 *Trans. Faraday Soc.* **33**, 1419.
Griffith, R. H. & Plant, J. H. G. 1939 Brit. Patent 529, 711.
Haul, R. & Schoon, T. 1939 *Z. Elektrochem.* **45**, 663.
Jellinek, K. & Zakowski, J. 1925 *Z. anorg. Chem.* **142**, 1.
Klemm, W. & Schüth, W. 1933 *Z. anorg. Chem.* **210**, 33.
Levi, G. R. & Baroni, A. 1935 *Z. Kristallogr.* **92**, 210.
Meade, E. M. 1938 *Chem. Age, N.Y.*, **39**, 300.
Mering, J. & Levialdi, A. 1941 *C.R. Acad. Sci., Paris*, **213**, 798.
Middleton, A. W. & Ward, A. M. 1935 *J. Chem. Soc.* p. 1459.
Selwood, P. W. 1946 *Chem. Rev.* **38**, 41.
Vanino, L. 1925 *Präparative Chemie*, Vol. **1**, 722, 3rd ed. Stuttgart.
Westgren, A. 1938 *Z. anorg. Chem.* **239**, 82.

The catalytic decomposition of simple heterocyclic compounds

II. Reaction kinetics and mechanism

By R. H. Griffith, J. D. F. Marsh and W. B. S. Newling
Fulham Laboratory, The Gas Light and Coke Co.

(*Communicated by E. J. Bowen, F.R.S.—Received* 9 *July* 1948—
Revised 31 *December* 1948—*Read* 17 *March* 1949)

The decomposition of thiophen, furane and pyrrole on molybdenum disulphide catalysts occurs by two-point adsorption, which leads to a half-hydrogenated state from which the hetero-atom is then broken off. Retardation is caused by molecules also capable of giving two-point adsorption. Tetrahydrothiophen and tetrahydrofurane are decomposed on the same catalysts, but the mechanism is probably simpler.

On nickel subsulphide catalysts, the tetrahydro-derivatives are decomposed much more easily than the unhydrogenated molecules, and they are held by single-point adsorption on the 111-plane of Ni_3S_2 on which pairs of hydrogen atoms are also adsorbed.

Thiophen and furane decompose by reaction with metallic nickel produced by reduction of Ni_3S_2. The nickel sulphide or oxide thus formed are subsequently reconverted to metal. Retardation of these changes arises from any reactant which can compete for nickel atoms, or which can cover the subsulphide surface and prevent its conversion to metal.

Introduction

The formation of tetrahydrofurane and of pyrrolidine is known to take place by hydrogenation of furane and pyrrole at normal pressures in the presence of metallic catalysts. With furane, Pd at 150° C (Starr & Hixon 1934), Ni at 180° C (Bourgignon 1908), and Os at 70° C (Shuikin & Chilikina 1936) are effective; with pyrrole, Rh at 100° C and Pd at 160° C (Zelinskii & Yurev 1931), and Ni at 180° C (Signaigo & Adkins 1936) are described as catalysts. Thiophen under similar conditions produces permanent poisoning of the catalyst (Roberti 1933). From the known properties of the catalysts which are mentioned above it may be concluded that the hydrogenation takes place by the same mechanism as that involved in saturation of the olefinic double bond (Twigg & Rideal 1940), where two-point adsorption of the molecule occurs. Breakdown of the heterocyclic ring does not set in under these conditions.

The dehydrogenation of pyrrolidine had also been observed at 360° C on Pt-asbestos (Wibaut, Molster, Kauffmann & Lenssen 1930) and on Pd-asbestos (Zelinskii & Yurev 1931) catalysts; tetra-hydrothiophen is converted to thiophen in contact with nickel-alumina at 350° C (Yurev & Borisov 1936). Although the tendency is therefore quite clearly in favour of the predominance of the unhydrogenated ring at high temperatures, it cannot be assumed that the hydrogenated structures are completely absent, so the suggestion of Moldavskii & Kumari (1934) that thiophen breaks down to give H_2S and a C_4 hydrocarbon through the intermediate formation of tetrahydrothiophen must be considered. These investigators, and also Cawley & Hall (1938), worked in static systems under pressure with catalysts consisting of the sulphides of molybdenum, nickel and cobalt; they demonstrated the presence of tetrahydrothiophen among the reaction products. In the present series of experiments at normal pressure, no tetrahydrothiophen could be detected under any conditions, whether the extent of thiophen decomposition was great or small. It seemed important, therefore, to compare the rate at which the hetero-atom was split off from the tetrahydro-compounds with that found for the unhydrogenated ring. As a result of these observations, the mechanism proposed by Moldavskii and Kumari proved untenable, and an alternative is now suggested.

THE DECOMPOSITION OF THIOPHEN, FURANE, PYRROLE AND THEIR TETRAHYDRO-DERIVATIVES BY HYDROGENATION

The preparation of the stocks of molybdenum sulphide catalysts is described in the previous paper. Nickel sulphide was prepared *in situ* from a stock of nickel hydroxide granules of similar size by treatment with purified hydrogen sulphide at 90° C, and reduced in pure hydrogen for 24 hr. at 400° C to give the nickel subsulphide catalysts used.

7·1 ml. of catalyst were heated electrically in a $\frac{3}{4}$ in. pyrex tube, the temperature being measured by a thermocouple.

The organic reagent was added to a stream of dry, oxygen-free hydrogen by a wick carburettor (Griffith & Hill 1938) modified so that it could be adjusted without interrupting the flow of gas. When using thiophen or tetrahydrothiophen, half the stream, which amounted to 14·2 l./hr., was passed through the reaction tube; the H_2S produced was removed and measured by aqueous $CdSO_4$, and the unchanged organic sulphur compound measured by combustion, washing out the SO_2 formed in a spiral washer, and periodic titration of the solution collected. The inlet concentration was determined simultaneously on the other half of the stream.

With the other reagents the course of the reaction was followed by measuring the water or ammonia formed. Calibration of the carburettor was carried out separately by adsorption of the reagent in weighed tubes of active carbon.

Retardants were added to the stream by a second carburettor, through a flow gauge or by a constant-head mercury displacement burette.

The kinetics of the hydrogenation of thiophen

The conversion of thiophen on Ni_3S_2 and MoO_2-MoS_2 was measured at three concentrations and various temperatures. Hydrogen was always present in large excess. Owing to a slow but appreciable fall in activity, corrections had to be made for the time factor in comparing experiments over long periods.

The corrected results are shown in figures 1 and 2. On Ni_3S_2 the reaction is nearly of the first order with respect to thiophen.

On MoO_2-MoS_2 the order is between 0·2 and 0·6 and increases as the temperature is raised or the concentration lowered.

Analysis of the products showed that mercaptans and tetrahydrothiophen were not present in detectable concentrations, and that the hydrocarbon product was approximately 60 % butene and 40 % butane.

On MoS_2 at 200° C the conversions were 69 % at 162 p.p.m., 31 % at 350 p.p.m. and 19 % at 690 p.p.m. These results show that the reaction is of approximately zero order with respect to thiophen.

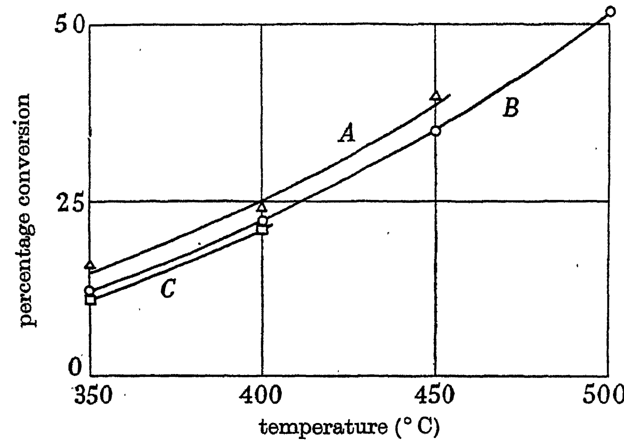


FIGURE 1. Thiophen hydrogenation on Ni_3S_2. Inlet concentrations: *A*, 206 p.p.m.; *B*, 378 p.p.m.; *C*, 550 p.p.m.

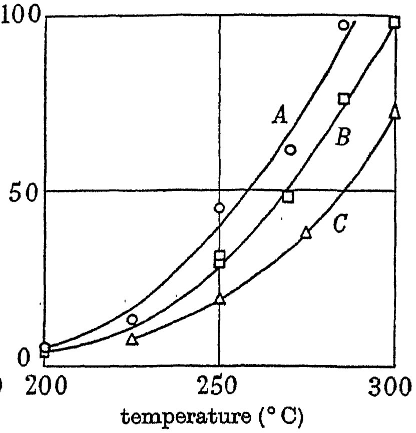


FIGURE 2. Thiophen hydrogenation on MoO_2-MoS_2 catalyst. Inlet concentrations: *A*, 155 p.p.m.; *B*, 275 p.p.m.; *C*, 500 p.p.m.

The retardation of thiophen hydrogenation by coal-gas constituents

Although the decomposition of thiophen occurs so readily on suitable catalysts in the presence of excess of pure hydrogen, the reaction does not take place in coal gas, which contains about 50 % of hydrogen. In order to identify the components responsible for this retardation effect, additions of the constituents of coal gas, at the concentrations of which they occur in coal gas, were made to the thiophen-hydrogen stream with thiophen at 250 p.p.m., and the mixture was passed over the MoO_2-MoS_2 catalyst at 300° C. The results are summarized in table 1.

TABLE 1. RETARDATION OF THIOPHEN HYDROGENATION

substance added	concentration	percentage of conversion in pure H_2
nil	—	100
ethylene	2 %	62
oxygen	1 %	84
methane	10 %	89
carbon monoxide	10 %	92
carbon disulphide	620 p.p.m.	57
sulphur dioxide	550 p.p.m.	76
hydrogen sulphide	550 p.p.m.	72
cyclopentadiene	710 p.p.m.	28

With all the retardants except cyclopentadiene the retardation was temporary, the percentage conversion returning to its initial value of about 70 % when the addition of retardant was stopped. With cyclopentadiene, however, the poisoning was permanent and progressive.

The retardation of thiophen hydrogenation on Ni_3S_2 *by* H_2S

On Ni_3S_2 at 400° C and at a concentration of 330 p.p.m. thiophen conversion was 40 %. Addition of 500 p.p.m. of H_2S reduced it to 20 % or to half of the unretarded conversion. The catalyst was cooled in the H_2-H_2S mixture and its magnetic mass susceptibility found to be $7{\cdot}5 \times 10^{-6}$ showing that all large groups of Ni atoms had been resulphided.

Comparison of various heterocyclic compounds

Measurements were made of the conversions of thiophen, tetrahydrothiophen, furane, tetrahydrofurane and pyrrole on MoO_2-MoS_2, MoS_2 and Ni_3S_2, in each case at one concentration and several temperatures. The results are shown in figures 3 and 4, and in table 2.

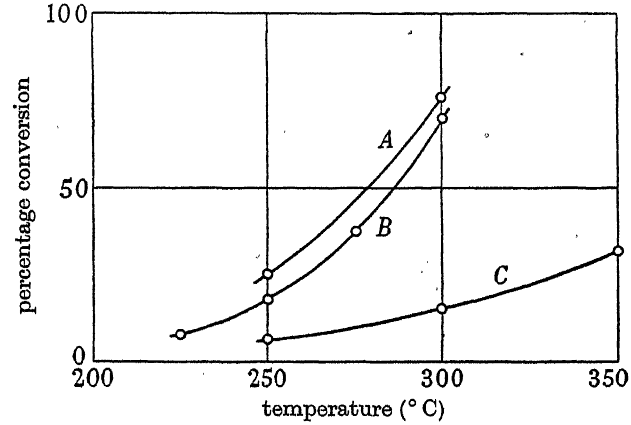


FIGURE 3. Conversion of heterocyclic compounds on MoO_2-MoS_2.

A, furane at 1090 p.p.m.; *B*, thiophen at 460 p.p.m.; *C*, pyrrole at 540 p.p.m.

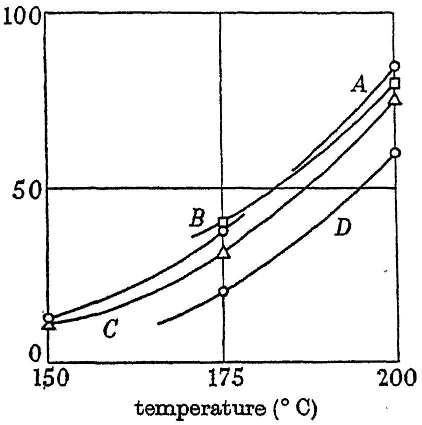


FIGURE 4. Conversion of heterocyclic compounds on MoS_2.

A, tetrahydrofurane at 895 p.p.m.; *B*, tetrahydrothiophen at 260 p.p.m.; *C*, furane at 1200 p.p.m.; *D*, thiophen at 260 p.p.m.

TABLE 2. CONVERSIONS OF HETEROCYCLIC COMPOUNDS ON Ni_3S_2

substance	initial concentration (p.p.m.)	percentage conversion		
		200° C	350° C	400° C
tetrahydrothiophen	280	41	100	100
tetrahydrofurane	870	0	25	55
thiophen	280	0	15	39
furane	1100	0	0	10
pyrrole	540	0	0	0

On molybdenum sulphide catalysts all the heterocylic compounds are reduced at nearly the same rates, though the tetrahydro-derivatives react somewhat faster than the unreduced compounds, and the oxygen compounds somewhat faster than their sulphur analogues, especially when allowance is made for differences in concentrations.

On Ni_3S_2, however, the reduced rings are much the more reactive. Thiophen and furane are partially decomposed at 350 to 400° C, but pyrrole is not attacked.

DISCUSSION OF RESULTS

(a) *The mechanism on molybdenum disulphide catalysts*

The information provided by the adsorption measurements, given in part I, is most significant. On MoS_2, hydrogen is very extensively adsorbed; at its maximum, about one molecule is held for each ten molecules of MoS_2, which is equivalent to about four molecules of hydrogen for every surface atom indicated by the low temperature adsorption of nitrogen. The sharp increase at 150° C to a value which is nearly constant at higher temperatures suggests that little penetration of the lattice is occurring, but that the hydrogen atoms can penetrate between the plates of the MoS_2 crystals where the nitrogen molecules are unable to do so.

On the mixed MoO_2-MoS_2 catalyst, hydrogen adsorption is much less, and presumably takes place almost entirely on such sulphide as is present, as it is known (Griffith & Hill 1935) that adsorption of hydrogen on MoO_2 is slight.

Thiophen, on the other hand, is more extensively adsorbed on the mixed catalyst than on the pure sulphide. As the MoS_2 catalyst is the more active, it is evident that the activated adsorption of hydrogen is the rate-controlling step. This conclusion is strongly supported by the fact that thiophen, furane and their tetrahydro-derivatives are all decomposed at about the same rate on the MoS_2 catalyst, although their adsorption isobars vary widely in character. It is significant that activated adsorption of hydrogen and the decomposition of the heterocyclic compounds both begin at about 150° C. Activated adsorption of the heterocyclic compounds also occurs in the temperature range where reaction begins. These facts suggest strongly that activated adsorption of both reactants is necessary.

In the decomposition of thiophen, there is no evidence for the intermediate formation of tetrahydrothiophen. It is suggested that the reaction proceeds by the following steps:

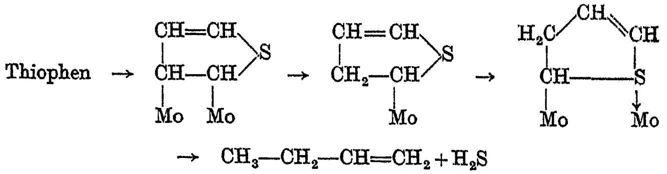
(1) Two-point adsorption of thiophen on a pair of adjacent Mo atoms in an exposed plane of the MoS_2 layer lattice.

(2) Conversion to the half-hydrogenated state by reaction with a hydrogen atom adsorbed on an adjacent Mo atom.

(3) Rupture of the carbon-sulphur bond by reaction of more hydrogen atoms with the half-hydrogenated molecule in which the sulphur atom has become linked to a molybdenum atom.

(4) Conversion of the adsorbed molecule to butene or butane and hydrogen sulphide by continued reaction with further hydrogen atoms.

These steps may be represented by the scheme:



In the decomposition of tetrahydrothiophen, a simpler mechanism is likely; this may involve a first step in which carbon and sulphur are held simultaneously on the molybdenum surface, but may only require single-point adsorption of the sulphur atom. Further attention is given to this suggestion in the discussion of the nickel sulphide catalysts.

The two-point adsorption postulated above is consistent with the known dimensions of the MoS_2 lattice and the length of the various bonds. It is very similar in character to that proposed by Herington & Rideal (1945) for unsaturated hydrocarbons on molybdenum oxide catalysts. In the molybdenum plane of MoS_2, the Mo atoms are 3·15 Å apart. Assuming the Mo-C bond to be 2·14 Å (Pauling 1944), two-point adsorption of thiophen by adjacent carbon atoms can occur if the Mo-C-C angle is expanded to 112° 5′ from the tetrahedral angle of 109° 28′. Similar arguments apply to furane or pyrrole.

This disposition of the organic molecule on the molybdenum surface still leaves ample space for the adsorption of hydrogen atoms on adjacent molybdenum atoms. The picture is in keeping with the recognized outstanding properties of molybdenum disulphide as a catalyst for the destructive hydrogenation of many types of organic substances.

It also explains the retardation of the decomposition of thiophen by cyclopentadiene, as this hydrocarbon has carbon-carbon distances practically identical with those in the heterocyclic molecule. It will therefore be capable of adsorption on similar surfaces and will ultimately block them completely by undergoing extensive polymerization.

(*b*) *The mechanism on nickel subsulphide catalysts*

The observations now recorded differ from those of earlier investigators in one important respect. This is that a streaming method has been used in place of the static system previously employed, and, as a result, the composition of the catalyst is different.

Nickel sulphide in a closed vessel may survive partly unchanged in contact with hydrogen and hydrogen sulphide. Under high pressure, it would be expected to catalyze the formation of tetrahydrothiophen by two-point adsorption of thiophen on atoms in one of those planes of the NiS lattice which contain only Ni atoms at a suitable spacing.

At normal pressures in a streaming system, unless the gas contains a high percentage of H_2S, no NiS survives and the catalyst is Ni_3S_2 with a completely different structure. Only in the 111-plane are Ni atoms suitably spaced to give two-point adsorption of the heterocyclic molecules or of hydrocarbons, but they occur in groups of three, as shown in figure 5, so that there is now no opportunity for the simultaneous presence of enough adsorbed hydrogen atoms to lead to reaction.

This feature is in agreement with the fact that Ni_3S_2 has no value as a catalyst for destructive hydrogenation of hydrocarbons, but is effective for the hydrogenation of reactants, such as CS_2, which are adsorbed at a single point by their sulphur atoms (Crawley & Griffith 1938). Tetrahydrothiophen falls into this category and is now shown to decompose readily on Ni_3S_2 at temperatures as low as 200° C, that is at approximately the lowest temperature at which activated adsorption of hydrogen is apparent. It is probable that a similar mechanism, involving only single-point adsorption on the catalyst, applies to the decomposition of the tetrahydro-derivatives on MoS_2 catalysts.

This mechanism cannot apply to the behaviour of thiophen and furane. It is suggested that these substances decompose by an entirely different cycle of changes, in which the solid does not function as a true catalyst. The first step is the production of metallic nickel by the reduction of Ni_3S_2. This is followed by reaction of the hetero-atom with the metal, the sulphide or oxide produced undergoing subsequent reduction in due course. This view is confirmed by the fact that H_2S and H_2O have now been found to be formed more quickly by the action of hydrogen at 400° C on Ni_2S_3 and NiO respectively than they are by the decomposition of thiophen or furane in contact with Ni_3S_2 catalysts. The heterocyclic components, moreover, react relatively slowly with preparations of the catalyst known to contain metallic nickel; when reaction does occur, it is appreciably faster with thiophen than with furane.

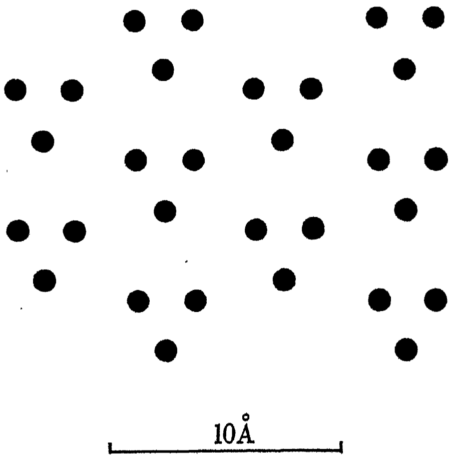


FIGURE 5. Ni atoms in the 111-plane of Ni_3S_2.

The retardation by H_2S of thiophen decomposition on Ni_3S_2 is in agreement with the proposed mechanism. Any other substance which is strongly adsorbed on the catalyst would also tend to suppress the production of metallic nickel at temperatures above 350° C, and it is therefore not difficult to understand the failure of this catalyst to decompose thiophen in coal gas.

REFERENCES

Bourgignon, A. 1908 *Chem. Zbl.* **1**, 1630.
Cawley, C. M. & Hall, C. C. 1943 *J. Soc. Chem. Ind., Lond.*, **62**, 116.
Crawley, B. & Griffith, R. H. 1938 *J. Chem. Soc.* pp. 720, 2034.
Griffith, R. H. & Hill, S. G. 1935 *Proc. Roy. Soc.* A, **148**, 194.
Griffith, R. H. & Hill, S. G. 1938 *J. Chem. Soc.* p. 2037.
Herington, E. F. G. & Rideal, E. K. 1945 *Proc. Roy. Soc.* A, **184**, 447.
Moldavskii, B. L. & Kumari, Z. 1934 *J. Gen. Chem. Moscow*, **4**, 298.
Pauling, L. 1944 *The nature of the chemical bond.* Cornell University Press.
Roberti, G. 1933 *Gazz. Chim. Ital.* **63**, 46.

Shuikin, N. I. & Chilikina, E. M. 1936 *J. Gen. Chem. Moscow*, **6**, 279.
Signaigo, F. K. & Adkins, H. 1936 *J. Amer. Chem. Soc.* **58**, 709.
Starr, D. & Hixon, R. M. 1934 *J. Amer. Chem. Soc.* **56**, 1595.
Twigg, G. H. & Rideal, E. K. 1940 *Trans. Faraday Soc.* **36**, 533.
Wibaut, J. P., Molster, C. C., Kauffmann, H. & Lenssen, A. M. 1930 *Rec. Trav. chim. Pays-Bas*, **49**, 1127.
Yurev, Y. K. & Borisov, A. E. 1936 *Ber. dtsch. chem. Ges.* **69**B, 1395.
Zelinskii, N. D. & Yurev, Y. K. 1931 *Ber. dtsch. chem. Ges.* **64**B, 101.

The hydrodynamical theory of film lubrication

By W. F. Cope,* *Engineering Division, National Physical Laboratory*

(*Communicated by Sir Charles Darwin, F.R.S.—Received* 3 *September* 1948)

The equations of film lubrication are derived from the general equations of hydrodynamics assuming only that the motion is steady, and that it takes place between two surfaces, in relative motion, which are both close together and nearly parallel. Consideration of the relative magnitude of the various terms under typical conditions as measured in a bearing shows that they are of very unequal orders of magnitude, and that in particular the inertia terms in the momentum and the dilatation and conductivity terms in the energy equation can be neglected. It is shown that film lubrication is possible if, and only if, either the distance between the surfaces decreases in the direction of motion (the geometric wedge), or the density of the fluid decreases in the same direction (the thermal wedge). These types of film are approximately equal when compared on a basis of equal film thickness and equal decrease. With the geometric wedge a much greater decrease, and therefore load-carrying capacity, is possible, but the thermal wedge from its simpler mechanical construction should be able to equalize matters by running with a thinner film.

The equations are reduced to non-dimensional form and the equation of state discussed. For the general case the integration is probably a major computing operation and, in view of the uncertainties in the exact form of the equation of state, not worth while. For the infinite bearing, on the other hand, integration is comparatively simple and has been carried out in the Mathematics Division, N.P.L., for a series of representative cases. The results show that viscosity variation has a profound effect on the performance of the thermal wedge, and that the additional wedging action provided by change of density is likely to be small unless the lubricating surfaces are close together. On the other hand, for surfaces close together and a small variation of viscosity the thermal wedge altogether outclasses the geometric.

INTRODUCTION

The analytical treatment of film lubrication was originally worked out by Reynolds (1886), on the basis of fluid flow through converging passages, and it was for long accepted that such passages were essential for film lubrication. Recently, Fogg (1946) reported the successful use of thrust bearings with parallel faces, and suggested as the explanation that the thermal expansion of the lubricant generated a thermal wedge. The analytical consequences of this have been partially investigated by Bower (1946) and Shaw (1947). In any case, the temperature distribution in a bearing is of some interest for its own sake, and though Christopherson (1941) and Cameron & Wood (1946) have studied it, the latter in connexion with the 'thermal wedge', no systematic modern treatment of the lubrication equations seems to exist. Since there are at least two different methods of generating lubricating films, it is desirable to examine the hydrodynamical theory of film lubrication *de novo*, to see if any other methods are possible. This paper derives the equations in a systematic manner and with the minimum assumptions for a fluid all of whose physical properties are variable.

---

* Principal Scientific Officer, Engineering Division, N.P.L.

NOTATION

So far as possible standard symbols have been selected, and fortunately it is possible, with a few exceptions, to use the corresponding capital letter when the equations are expressed in reduced non-dimensional form. The most important exception is viscosity, for which the standard symbol is η, but H cannot be used for the reduced form η/η_1, so Λ has been selected on the ground that λ is used for viscosity in much of the literature of the subject. The suffix 0 refers to point of maximum pressure, suffix 1 refers to inlet conditions and suffix 2 to conditions at the point of closest approach. In full the notation is:

x, y, z	Cartesian co-ordinates referred to a rectangular system.	
u, v, w	velocities in directions of x, y and z respectively.	
p	pressure	of the fluid expressed in any convenient self consistent system. The same unit is used for work and heat thereby avoiding the explicit appearance of J.
ρ	density	
T	temperature	
η	viscosity	
k	thermal conductivity	
C	specific heat	
E	internal energy; assumed to be a function of T only and therefore given by $E = CT$ with C constant.	
Φ	dissipation function; the rate (per unit of volume) at which work is being done against viscosity and therefore the rate at which mechanical energy is disappearing.	
α	thermal coefficient of cubical expansion.	
h	film thickness.	
U	velocity of moving surface.	
X, Y, Z	body force per unit volume in equations (1b) and (2) only.	
Δ	dilatation.	
K	the ratio, inlet film thickness to 'effective film thickness' at point of closest approach (defined by equation (11)).	

The substitutions bringing the equations to reduced form are:

$$P = p/(6\eta_1 U/l), \qquad R = \rho/\rho_1, \qquad H = h/l, \qquad X = x/l,$$

$$Y = y/l, \qquad \Lambda = \eta/\eta_1, \qquad \Theta = T/(\eta_1 U/\rho_1 Cl).$$

l is any representative length and is taken as the length of bearing in the x-direction in the body of the paper.

Conductivity effects are usually negligible, but the form for the 'reduced conductivity' is $k/(\rho_1 CUl)$. It is given both for completeness, and because it is identical with the 'Reynold's thermal number' of Bisacre & Bisacre (1947). The reason for the identity is, simply, that in both cases the process is one of deforming a substance and generating heat by what may be termed a shearing action, so that the analysis is concerned with the ratio between the rate at which this heat is conducted through the substance and the rate at which it is generated.

It would be very convenient in many ways to work with a reduced temperature proportional to $T-T_1$, but it will be found that this introduces difficulties in the mode of expression of the law of viscosity variation with temperature, for instance $\Lambda \propto T^{\beta}$ implies $\Lambda \propto e^{\beta(T-T_1)}$, and a consideration of the orders of magnitude of β and T confirms what is otherwise known from physical measurements, that to expand the exponential and retain only the first power of $T-T_1\{\Lambda \propto 1+\beta(T-T_1)\}$ is unlikely to be sufficiently accurate.

Precisely the same set of equations result from the substitutions

$$P = p/(6\eta_1 Ul/h_2^2), \quad H = h/h_2, \quad \Theta = T/(\eta_1 Ul/\rho_1 C h_2^2),$$

the remaining substitutions being as before.

The substitution is the logical extension of that used by Christopherson (1941), but it has the drawback of introducing a quantity h_2, which is not known *a priori*, and this is a disadvantage from the point of view of actual design. Moreover, there are two 'standard' lengths h_2 and l involved which can be a source of confusion.

The definitions of the several criteria are:

$$\text{Load criterion } \Delta_L = \frac{\text{load per unit area}}{\eta_1 U/l}\frac{h_2^2}{l^2} = 6H_2^2\int_0^1 P\,dX.$$

$$\text{Drag criterion } \Delta_D = \frac{\text{frictional drag per unit area}}{\eta_1 U/l}\frac{h_2}{l} = 4H_2\int_0^1 \frac{1}{H}\left(1-\frac{3}{4}\frac{R_0 H_0}{RH}\right)dX.$$

$$\text{Friction criterion } \Delta_\mu = \Delta_D/\Delta_L.$$

$$\text{Temperature criterion } \Delta_T = \frac{\text{temperature rise through bearing}}{\lambda_1 U/\rho_1 Cl}\,h_2^2 = (\Theta_2-\Theta_1)H_2^2.$$

Derivation of equations

The fundamental equations of hydrodynamics when all the physical properties of the fluid are variables can be compactly expressed in vector notation (Cope 1942). But since, in the sequel, they are wanted in rectangular Cartesian co-ordinates, it is best to give them at once in that form. In the steady state, when the flow is independent of time, they are:

The continuity equation (1a) expressing constancy of mass.

The momentum equations (1b) expressing Newton's second law, sometimes called the Navier-Stokes equations.

The energy equation (1c) expressing constancy of energy.

$$\left.\begin{aligned}
&\frac{\partial}{\partial x}(\rho u)+\frac{\partial}{\partial y}(\rho v)+\frac{\partial}{\partial z}(\rho w) = 0, && (a)\\
&\rho u\frac{\partial u}{\partial x}+\rho v\frac{\partial u}{\partial y}+\rho w\frac{\partial u}{\partial z} = \rho X-\frac{\partial p}{\partial x}+\eta\nabla^2 u\\
&\quad +\tfrac{1}{3}\eta\frac{\partial \Delta}{\partial x}+\frac{4}{3}\frac{\partial \eta}{\partial x}\frac{\partial u}{\partial x}+\frac{\partial \eta}{\partial y}\frac{\partial u}{\partial y}+\frac{\partial \eta}{\partial z}\frac{\partial u}{\partial z}+\frac{\partial \eta}{\partial y}\frac{\partial v}{\partial x}+\frac{\partial \eta}{\partial z}\frac{\partial w}{\partial x}-\frac{2}{3}\left(\frac{\partial \eta}{\partial x}\frac{\partial v}{\partial y}+\frac{\partial \eta}{\partial y}\frac{\partial w}{\partial z}\right) = 0, && (b)\\
&\rho u\frac{\partial E}{\partial x}+\rho v\frac{\partial E}{\partial y}+\rho w\frac{\partial E}{\partial z}+p\Delta = \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right)+\frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right)+\Phi. && (c)
\end{aligned}\right\}\quad (1)$$

Equation (1*b*) expresses the rate of change of momentum in the direction of x, and there are, of course, two similar equations for the rates of change in the directions of y and of z respectively.

In equation (1*c*) the first group of terms represents, for any element of the fluid, the rate of decrease of internal energy, the second the work done by the pressures in compressing the fluid (the dilatation term), the third the rate at which energy in the form of heat is being conducted away, and the fourth the rate at which work is being done against viscosity (the dissipation function). It should be particularly noted that the only assumption about the nature of the fluid made in deriving the equation is that E is a function of T only.

The analytical expression for Φ is

$$\Phi = \eta\left[2\left\{\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial v}{\partial y}\right)^2+\left(\frac{\partial w}{\partial z}\right)^2\right\}+\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial v}{\partial z}+\frac{\partial w}{\partial y}\right)^2+\left(\frac{\partial w}{\partial x}+\frac{\partial u}{\partial z}\right)^2-\tfrac{2}{3}\Delta^2\right]. \quad (2)$$

The essential assumption which leads to the hydrodynamical equations of film lubrication is that the flow is confined to a thin film. More precisely, it is assumed that the flow is confined between two surfaces that are (1) either parallel or but slightly inclined; (2) close together; and (3) in relative tangential motion, and the method employed is akin to that used in deriving the boundary-layer equations in fluid dynamics (see, for instance, Goldstein 1938, Ch. IV). The arguments are essentially physical, though their conclusions are expressed in analytical form and amount to saying that, because of the configuration envisaged, certain quantities cannot vary much in certain directions and that certain quantities are far more important than others. In the first place the assumption that the surfaces are close together implies that if the surfaces are curved their radius of curvature is very much greater than their distance apart. In other words, the distinction between polar and rectangular co-ordinates can be neglected and the surfaces treated as planes, and in anticipation of this conclusion equations (1) and (2) were given in rectangular Cartesian form. In the applications which follow one plane is regarded as being at rest, the origin is taken at the beginning of the film, the axis of z perpendicular to it, the plane $z = 0$ as the moving plane and the axis of x in the direction of motion.

Since the surfaces are close together, nearly parallel and in relative tangential motion it is reasonable to assume:

(1) That the variation of pressure, density, temperature, viscosity and thermal conductivity across the film are far less important than their variation along it. That is, it is assumed that

$$p,\ \rho,\ T,\ \eta \text{ and } k \text{ are functions of } x \text{ and } y \text{ only.}$$

(2) That the velocity gradients across the film are much more important than velocity gradients parallel to it. That is, it is assumed that

$$\frac{\partial u}{\partial z} \gg \frac{\partial u}{\partial x}, \quad \frac{\partial^2 u}{\partial z^2} \gg \frac{\partial^2 u}{\partial z\,\partial x} \gg \frac{\partial^2 u}{\partial x^2},$$

so that only the first named need be retained in groups of such terms.

(3) That the fluid velocity perpendicular to the film is very small. That is, it is assumed that

$$w = 0.$$

It is also assumed that the body forces, including the centrifugal force due to rotation, can be neglected; that is,

$$X = Y = Z = 0,$$

and that the specific heat is constant; that is,

$$E = CT.$$

When the assumptions have been applied to equations (1) they become

$$\left.\begin{aligned}
&\frac{\partial}{\partial x}(\rho u)+\frac{\partial}{\partial y}(\rho v) = 0, && (a)\\
&\left.\begin{aligned}
&\rho u\frac{\partial u}{\partial x}+\rho v\frac{\partial u}{\partial y}+\frac{\partial p}{\partial x} = \eta\frac{\partial^2 u}{\partial z^2},\\
&\rho u\frac{\partial v}{\partial x}+\rho v\frac{\partial v}{\partial y}+\frac{\partial p}{\partial y} = \eta\frac{\partial^2 v}{\partial z^2},
\end{aligned}\right\} && (b)\\
&\rho Cu\frac{\partial T}{\partial x}+\rho Cv\frac{\partial T}{\partial y}+p\Delta = \frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right)+\eta\left\{\left(\frac{\partial u}{\partial z}\right)^2+\left(\frac{\partial v}{\partial z}\right)^2\right\}, && (c)
\end{aligned}\right\} \quad (3)$$

The third momentum equation, to the same approximation, becomes

$$\frac{\partial \eta}{\partial x}\frac{\partial u}{\partial z}+\frac{\partial \eta}{\partial y}\frac{\partial v}{\partial z}\left(= -\frac{\partial p}{\partial z}\right) = 0, \qquad (3d)$$

so that $\partial p/\partial z$ is at most $O(\partial u/\partial z)$, whereas $\partial p/\partial x$ and $\partial p/\partial y$ are $O(\partial^2 u/\partial z^2)$; this verifies, *a posteriori*, the approximate constancy of p across the film. Equation ($3d$) therefore merely states that η is a function of u and v; this is already implicit in equations (3) and no new restriction is imposed. If it ever becomes necessary to take the variation of p with z into account, then the integral of equation ($3d$) can be used (Christopherson 1941) to correct values of p obtained on the assumption that p is independent of z.

The second and final step in the derivation process is to see if any further simplification can be effected by considering the relative magnitudes of the various terms. This is done by taking from actual measurements for a typical set of conditions values of ρ, p, h, U, etc., calculating from them the orders of magnitude of the several terms in equations (3) and then neglecting any terms which are thereby shown to be unimportant. Table 1 gives, in the absolute c.g.s. system, what are thought to be typical figures for both gas and liquid lubricated bearings. Furthermore, since $\partial u/\partial z$ is of the same order as U/h, from table 1 it can be inferred that $\partial u/\partial z$ for a gas is 10^8, $\partial^2 u/\partial z^2$ is 10^{12} and so on. Proceeding in this way table 2 is compiled.

This shows that in equation ($3b$) the inertia terms ($\rho u(\partial u/\partial x)$, etc.) are about 10^{-4} of the viscous terms and can therefore be neglected. The Reynolds number is comparatively high, and it is interesting to speculate whether inertial effects lead to turbulence in the film before it becomes necessary to include the inertia terms in ($3b$). Similarly, in ($3c$) the dilatation terms are 10^{-3} of the dissipation terms even for a gas, and the conductivity terms are the space derivates of quantities of the order 10^{-12} of the dissipation terms. So that, as very rapid changes of temperature gradient are unlikely, both these groups can be neglected.

TABLE 1

	ρ	p	h	η	U	k	C	$\partial u/\partial x$
gas	10^{-3}	10^{7}	10^{-4}	10^{-4}	10^{4}	10^{-4}	10^{7}	10^{2}
liquid	1	10^{8}	10^{-3}	10^{-1}	10^{4}	10^{-3}	10^{7}	10^{2}

TABLE 2

	R	$\rho u(\partial u/\partial x)$	$\eta(\partial^2 u/\partial z^2)$	$\eta(\partial u/\partial z)^2$	$p\Delta$	$k(\partial T/\partial x)$	$\partial p/\partial x$
gas	10	10^{3}	10^{8}	10^{12}	10^{9}	10^{-1}	10^{8}
liquid	100	10^{6}	10^{9}	10^{15}	10^{10}	10	10^{9}

It is implicit in assumptions (1) that heat transfer between the lubricant and the bearing surface is negligible. This implies both (*a*) that all the heat generated by friction is carried away by the lubricant and (*b*) that there is not thermal circulation by which heat leaves the lubricant at points of high temperature and is conducted back through the bearing to be reintroduced at points of low temperature.

As regards (*a*), it can be shown that the heat loss from the bearing to its surroundings by natural convection is of the order of 10^{-1}. So that for an appreciable quantity of heat to be abstracted in this way the area of the surfaces (pedestal, shaft, etc.) must be a billion (10^{12}) times the volume of the film. This seems very unlikely, and so the assumption that all the heat generated is carried away by the lubricant seems reasonable.

As regards (*b*), the conductivity of the bearing material will be about a thousand times that of the lubricant. On the other hand most lubricants have small forced convection coefficients (the difficulty of designing a good oil cooler is well known) and thermal conductivity effects within the body of the lubricant are negligible. So it seems likely that thermal circulation is only important (if at all) in rather exceptional circumstances and that even then its effects would be purely local.

Some of these conclusions are intuitively obvious, but the method employed has the great advantage of indicating not only what terms can be neglected, but also of giving the order of error that is made in neglecting them. Moreover, if a case that may be exceptional is encountered, it is only a few minutes' work to amend tables 1 and 2 to suit the altered conditions and to establish the same or another system of approximation.

The equations now become

$$\left.\begin{aligned}
&\frac{\partial}{\partial x}(\rho u)+\frac{\partial}{\partial y}(\rho v)=0, && (a)\\
&\left.\begin{aligned}\frac{\partial p}{\partial x}&=\eta\frac{\partial^2 u}{\partial z^2},\\ \frac{\partial p}{\partial y}&=\eta\frac{\partial^2 v}{\partial z^2},\end{aligned}\right\} && (b)\\
&\rho C\left(u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}\right)=\eta\left\{\left(\frac{\partial u}{\partial z}\right)^2+\left(\frac{\partial v}{\partial z}\right)^2\right\}. && (c)
\end{aligned}\right\} \qquad (4)$$

The standard form of manipulation, including the application of the velocity boundary conditions

$$\left.\begin{aligned} z = 0,\ u = U,\ v &= 0 \\ z = h,\quad u = v &= 0 \end{aligned}\right\},$$

eliminates u and v from these equations and they become

$$\left.\begin{aligned} &6U\left[\frac{\partial}{\partial x}\left\{\rho h\left(1-\frac{h^2}{6\eta U}\frac{\partial p}{\partial x}\right)-\frac{\partial}{\partial y}\left(\frac{\rho h^3}{6\eta U}\frac{\partial p}{\partial y}\right)\right\}\right] = 0, && (a)\\ &6U\rho Ch\left[\left(1-\frac{h^2}{6\eta U}\frac{\partial p}{\partial x}\right)\frac{\partial T}{\partial x}-\frac{h^2}{6\eta U}\frac{\partial p}{\partial y}\frac{\partial T}{\partial y}\right] = \frac{12\eta U^2}{h}\left[1+\frac{h^4}{12\eta^2U^2}\left\{\left(\frac{\partial p}{\partial x}\right)^2+\left(\frac{\partial p}{\partial y}\right)^2\right\}\right]. && (b) \end{aligned}\right\} \tag{5}$$

Equation (5 a) reads as if viscosity variation had been neglected in obtaining it from (4a) and (4b). The justification of this is that the integrals of (4a) and (4b), with viscosity variation in the z direction taken into account, are

$$u = \frac{h-z}{h}U - \frac{z(h-z)}{\eta_0}\frac{\partial p}{\partial x} - \frac{z}{\eta_0}\int_0^h\left(u\frac{\partial \eta}{\partial z}\right)dz, \tag{6}$$

and a similar equation in v. Assumptions (1) can be used to justify neglecting $\partial\eta/\partial z$, so the last term is dropped and the equation reduces to its familiar form.

The assumptions which have led to equations (5) have been made on physical grounds. But an examination of the analytical consequences of a less restrictive set suggests that they are the most general that will still permit the application of the velocity boundary conditions to obtain equations (5) from equations (4); except only that provision could be made, through (3d), for variation of pressure through the film. Apart from this possible enlargement of the conditions, if a less restrictive set of assumptions were ever needed, it would be necessary to solve equations (4) as they stand. This would certainly prove both longer and more difficult, because not only would there be three equations in the set instead of two, but all three would be rendered more complicated because of the inclusion of hitherto neglected terms.

Equations (5) can be further simplified and made much more convenient for numerical work by reducing them to non-dimensional form. In terms of the non-dimensional or reduced variables they are

$$\left.\begin{aligned} &\frac{\partial}{\partial X}\left\{RH\left(1-\frac{H^2}{\Lambda}\frac{\partial P}{\partial X}\right)-\frac{\partial}{\partial Y}\left(\frac{RH^3}{\Lambda}\frac{\partial P}{\partial Y}\right)\right\} = 0, && (a)\\ &RH\left(1-\frac{H^2}{\Lambda}\frac{\partial P}{\partial X}\right)\frac{\partial \Theta}{\partial X}-\frac{RH^3}{\Lambda}\frac{\partial P}{\partial Y}\frac{\partial \Theta}{\partial Y} = 2\frac{\Lambda}{H}\left[1+\frac{3H^4}{\Lambda^2}\left\{\left(\frac{\partial P}{\partial X}\right)^2+\left(\frac{\partial P}{\partial Y}\right)^2\right\}\right]. && (b) \end{aligned}\right\} \tag{7}$$

Equation of state

In equations (7), Λ is a known function of P and Θ, and H a known function of X and Y. So one more equation is needed as there are three unknowns (P, R and Θ), and this is provided by the relation between P, R and Θ called the equation of state. For lubrication with a gas film the simplest assumption, that the gas obeys the perfect gas law, will be adequate, and since P and Θ are the quantities of practical importance the procedure might be to eliminate R from (7a) and (7b) by the substitution

$P = \frac{1}{6}(\gamma - 1) R\Theta$. For lubrication with a liquid film the choice of an equation of state is more difficult. Of recent years there have been many attempts to derive an equation of state for liquids (e.g. *J. Chem. Phys.* 1937 onwards), but a perusal of the extensive literature of the subject leaves the impression that it can be derived with any confidence only with simple liquids which excludes lubricating oils. Moreover, even for simple liquids the equations proposed are modified van der Waals type of considerable algebraic complication, so that their introduction into equations (7) would increase the complications to such an extent that the solution would probably be a major computing operation. At present no guarantee could be given that the equation of state was above criticism as an analytical representation of the properties of the liquid, so it appears unjustifiable to embark on this work.

If it be sufficiently accurate to ignore the variation of R, Λ, etc., with temperature, or if their variation with temperature can be replaced by a variation, known *a priori*, with X and Y, then (7*a*) becomes an equation for P only. And the solution obtained can be inserted into (7*b*) which then becomes an equation for Θ only, but in practice this last step can be omitted when no interest attaches to the temperature difference as such. This situation, which also arises in other branches of applied mechanics, enables the equations to be solved successively instead of simultaneously, but only in the order (7*a*) to (7*b*). Although, fortunately, this is the order of practical importance, solution in the reverse order is always impossible, and as soon as it becomes necessary to take variation with temperature into account the equations become interlocked and must be solved simultaneously.

Comparison with previous equations

With ρ constant (5*a*) reduces to the classical Reynolds (1886) form, and with ρ variable and h constant it reduces to the form used by Cameron & Wood (1946).

The energy equation (5*b*) is new and does not agree with the equation used by Christopherson (1941) and Cameron & Wood (1946) in calculating temperature rise. Both these authors use an equation which is (in the present notation)

$$6U\rho Ch\left\{\left(1-\frac{h^2}{6\eta U}\frac{\partial p}{\partial x}\right)\frac{\partial T}{\partial x}-\frac{h^2}{6\eta U}\frac{\partial p}{\partial y}\frac{\partial T}{\partial y}\right\} = 12\frac{\eta U^2}{h}\left[1+\frac{h^2}{2\eta U}\frac{\partial p}{\partial x}\right], \tag{8}$$

and differs in the last term of the right-hand side. An examination of the orders of the terms inside the square bracket shows that in (8) they are (for a liquid) 10^8 and 10^3 and in (5*b*) 10^8 and 10^5 respectively. The U^2 is thus the dominant term, but equation (8) considerably underestimates the effect of pressure gradient. Results from (5*a*) and (8) in combination would be expected to be quite reliable as long as the pressure gradient was not too great but to diverge more and more as the latter increased. This point has been verified (in Mathematics Division, N.P.L.) by integrating (5*a*) and (5*b*) for the two extreme conditions assumed by Cameron & Wood (1946) and comparing the results with the integration of (5*a*) and (8). The two solutions are, for practical purposes, identical at the lightly loaded extreme but appreciably different at the other, with indications that the divergence was beginning to increase rapidly. In any case it is obvious that since $\partial u/\partial z$ involves $\partial p/\partial x$, the energy equation must contain terms in $(\partial p/\partial x)^2$.

MECHANISM OF FILM FORMATION

The understanding of the mechanism of film formation is made much easier by neglecting all variation with Y, that is, by considering an infinite bearing, and for clarity the argument, though perfectly general, will be worded so as to apply directly to thrust bearings provided the speed is such that centrifugal forces are unimportant. This particular case has been selected because the question of geometrical divergence of the film does not arise, and so the boundary conditions can be stated in a form which facilitates solution.

In these circumstances equations (7) become

$$\left.\begin{aligned}&\frac{d}{dX}\left\{RH\left(1-\frac{H^2}{\Lambda}\frac{dP}{dX}\right)\right\} = 0, &&(a)\\ &RH\left(1-\frac{H^2}{\Lambda}\frac{dP}{dX}\right)\frac{d\Theta}{dX}-\frac{2\Lambda}{H}\left\{1+\frac{3H^4}{\Lambda^2}\left(\frac{dP}{dX}\right)^2\right\} = 0. &&(b)\end{aligned}\right\}\tag{9}$$

($9a$) can be integrated once and the equations then become

$$\left.\begin{aligned}&\frac{dP}{dX} = \frac{\Lambda}{H^2}\left(1-\frac{R_0H_0}{RH}\right), &&(a)\\ &\frac{d\Theta}{dX} = \frac{8}{R_0H_0}\frac{\Lambda}{H}\left\{1-\frac{3}{2}\frac{R_0H_0}{RH}+\frac{3}{4}\left(\frac{R_0H_0}{RH}\right)^2\right\}, &&(b)\end{aligned}\right\}\tag{10}$$

where R_0H_0 is the constant of integration.

Now for successful film lubrication the pressure must rise to a maximum and fall again, that is, dP/dX must vanish; therefore equation ($10a$) has been written in a form which emphasizes that R_0 and H_0 are the values of R and of H where dP/dX vanishes. The equation shows that evanescence of dP/dX cannot be achieved by variation of Λ alone, since viscosity is in practice a monotonic function, but that it can be achieved by suitable decrease of R, of H or of their product. Qualitatively the effect is the same for variation of R or of H, but quantitatively it will be different, since $1/H^2$ occurs outside the bracket; a decrease of H will produce a greater total change in dP/dX, so varying H is likely to be more effective than varying R. If R and H both vary in such a way that RH is always $> R_0H_0$ then again film lubrication is impossible.

Now the pressure in a bearing increases manyfold, but the temperature only a fewfold, so that:

For a gas R ($\propto P/\Theta$) will increase. So it seems unlikely that it will be possible to run parallel face bearings with gas as a lubricant, and it will be noticed that it seems likely that the variation of H will have to be considerable to counterbalance the increase in R.

For a liquid the increase of R with P is usually small compared with its decrease with Θ, and the overall variation of R may prove sufficient in some cases for successful film lubrication of parallel face bearings.

Finally, since for all real fluids R and P increase or decrease together, increase of pressure tends to inhibit the formation of a film. It seems then that the only suitable variation of R is the decrease with temperature, and therefore, that the description of the mechanism of film lubrication of parallel face bearings as due to a 'thermal wedge' is fully justified. The other type, with H varying, might well be called the 'geometric wedge'.

Solutions of the equations

Equations (7) are complicated, and obtaining solutions even with the aid of modern computing techniques is likely to be a lengthy process. In addition, an equation of state of some kind is needed, and this at the moment can only be given with confidence for gases. Probably therefore the trouble of obtaining solutions is not worth while except where numerical results are urgently needed for a special application for which an *ad hoc* equation of state is known to be satisfactory.

On the other hand, equations (10) (repeated below for convenience)

$$\left.\begin{aligned} \frac{dP}{dX} &= \frac{\Lambda}{H^2}\left(1-\frac{R_0H_0}{RH}\right), && (a)\\ \frac{d\Theta}{dX} &= \frac{8}{R_0H_0}\frac{\Lambda}{H}\left\{1-\frac{3}{2}\frac{R_0H_0}{RH}+\frac{3}{4}\left(\frac{R_0H_0}{RH}\right)^2\right\}, && (b)\end{aligned}\right\} \qquad (10)$$

with boundary conditions

$$X=0,\quad P=0,\quad \Theta=\Theta_1,\quad R=\Lambda=1;\quad X=1,\quad P=0,$$

and laws of variation of H, R and Λ

$$\frac{H}{H_2}=H_1\left(1-\frac{H_1-H_2}{H_1}X\right),\quad R=1-A(\Theta-\Theta_1),\quad \Lambda=(\Theta/\Theta_1)^\beta,$$

where H_1, H_2, A and β are known *a priori*, are comparatively easy to solve. The reason is that (10*b*), with R_0H_0 supposed known, can be integrated immediately to give the dependence of Θ on X. From this relationship the dependence of R and Λ on Θ can be replaced by a dependence on X so that (10*a*) becomes integrable; if the supposed value of R_0H_0 is correct then P will vanish for $X=1$; if it does not, then another value of R_0H_0 is taken and the process repeated as often as is necessary to make P vanish for $X=1$ to the required order of accuracy. In use the third assumption is almost always sufficiently accurate, and with practice the proper value of R_0H_0 can be guessed second or sometimes even first time.

Solutions obtained in this way from equations (10) cannot of course pretend to represent all that goes on in an actual bearing. Besides any approximation involved in the assumptions which lead to equations (7), side leakage has been neglected, the simplest possible equation of state is used, and the law of viscosity variation selected is a compromise between the viscosity variation that is actually obtained and the necessity of having a law of variation that is analytically simple. Nevertheless, it seems that information of real value as indicating trends could be obtained or inferred from a set of solutions with H_1, H_2, A and β selected to cover the range of variation actually encountered in practice.

The solutions of (10*a*) for Λ and R constant and H a linear function of X are classical (Lamb 1924). Harrison (1912) has studied it for the case Λ constant and

$R \propto P$ in analyzing Kingsbury's (1897) measurements with air as a lubricant, Bower (1946) has discussed it when one or more of Λ, R and H are linear functions of X, and Shaw (1947) for R a linear function of X. It seems worth while, therefore, to obtain or quote in algebraic form the solutions for the two simplest cases, namely, H a linear function of X (the pure geometric wedge), and R a linear function of Θ (the pure thermal wedge). The work is given below in parallel columns, and because of its complication the solution of the energy equation is not given. This solution does, however, show that, for constant viscosity and light loads at any rate, Θ is approximately a linear function of X. This means that the analysis can be considerably simplified by replacing the linear dependence of R on Θ by a linear dependence of R on X; accordingly the formulae for the thermal wedge in the table below have been derived for $R \propto X$. Both Bower (1946) and Shaw (1947) introduced this proportionality as an *a priori* assumption.

TABLE 3

	geometric wedge	thermal wedge
equations	$\dfrac{dP}{dX} = \dfrac{1}{H^2}\left(1-\dfrac{H_0}{H}\right)$ $\dfrac{d\Theta}{dX} = \dfrac{8}{H_0}\dfrac{1}{H}\left\{1-\dfrac{3}{2}\dfrac{H_0}{H}+\dfrac{3}{4}\left(\dfrac{H_0}{H}\right)^2\right\}$	$\dfrac{dP}{dX} = \dfrac{1}{H_2^2}\left(1-\dfrac{R_0}{R}\right)$ $\dfrac{d\Theta}{dX} = \dfrac{8}{H_2^2}\left\{1-\dfrac{3}{2}\dfrac{R_0}{R}+\dfrac{3}{4}\left(\dfrac{R_0}{R}\right)^2\right\}$
boundary conditions	$X = 0, H = H_1, P = 0, \Theta = \Theta_1$ $X = 1, H = H_2, P = 0, \Theta = \Theta_2$	$X = 0, R = 1, P = 0, \Theta = \Theta_1$ $X = 1, R = R_2, P = 0, \Theta = \Theta_2$
def. of K	H_1/H_2	$1/R_2$
law of variation	$H = KH_2\left(1-\dfrac{K-1}{K}X\right)$	$R = 1-\dfrac{K-1}{K}X$
H_0 or R_0	$2KH_2/(K+1)$	$(K-1)/K\ln K$
P	$\dfrac{(H_1-H)(H-H_2)}{(K^2-1)H_2^2H^2}$	$\left\{\dfrac{K}{K-1}(1-R)-\dfrac{\ln(1/R)}{\ln K}\right\}\Big/H_2^2$
Δ_L	$\dfrac{6}{(K-1)^2}\left\{\ln K-2\dfrac{K-1}{K+1}\right\}$	$3\left(\dfrac{K+1}{K-1}-\dfrac{2}{\ln K}\right)$
Δ_D	$\dfrac{2}{K-1}\left\{2\ln K-3\dfrac{K-1}{K+1}\right\}$	1
Δ_T	$\dfrac{2}{K}\left\{2\dfrac{K+1}{K-1}\ln K-3\right\}$	$6\dfrac{(K-1)^2}{K(\ln K)^2}-4$

The integration of equations (10) with their boundary conditions and laws of variation given on p. 210 has been carried out in Mathematics Division, N.P.L.; the results are given in table 3 and plotted as figures 1 to 5. The overall accuracy of the results is about 1 %, which is ample in relation to the assumptions on which the analysis is based. Typical values have been taken for U, l, η_1, etc., and the only points needing comment are the values selected for A, β, and the proper definition of K when both geometric and thermal effects are present. Two values of A have been used; zero corresponding to no change of density with temperature, and a value corresponding to a coefficient of cubical expansion of 10^{-3} which seems to be typical of many fluids. The three values of β are zero for constant viscosity, and $-1\frac{1}{2}$ and

– 3 which are the sort of law of variation of the 'light' and 'heavy' lubricating oils respectively; roughly speaking they imply that the viscosity at 100° C is 1/10 and 1/100 of its value at 20° C.

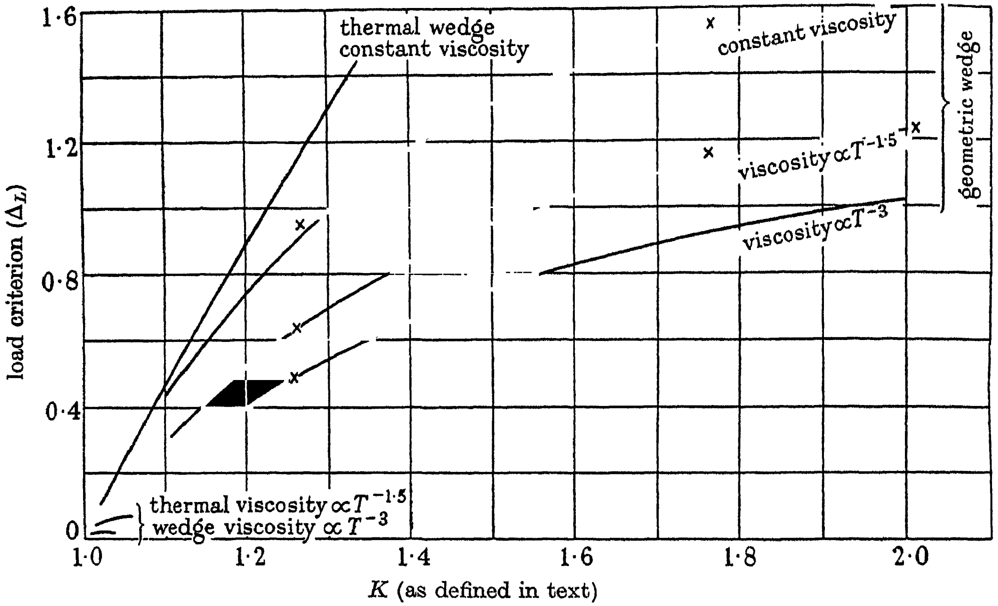


FIGURE 1. × = geometric wedge with allowance for oil expansion, $\alpha = 10^{-3}$.

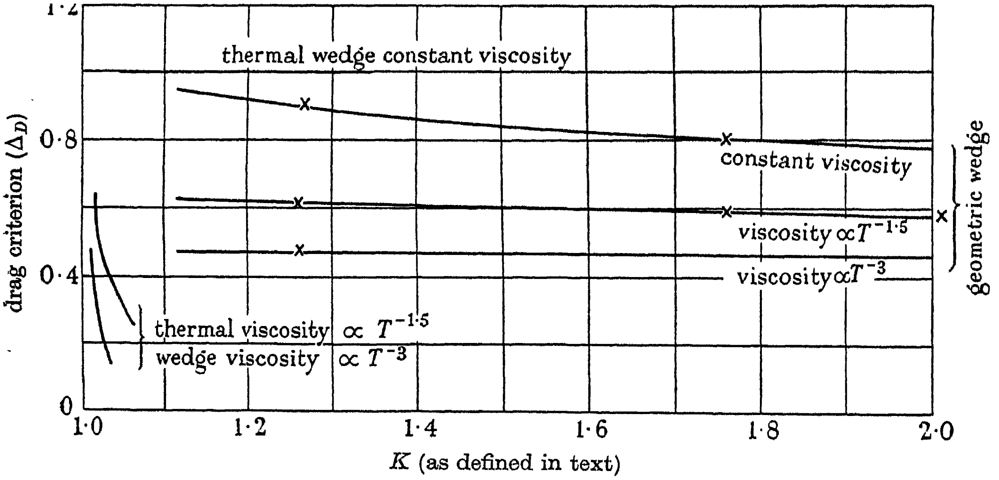


FIGURE 2. × = as for figure 1.

The definition of K also requires clarification when both geometric and thermal wedges are present. The preceding discussion shows that $1/R_2$ plays the same part for the thermal wedge as H_1/H_2 does for the geometric. The simplest procedure is to define K as the ratio $\dfrac{\text{inlet film thickness}}{\text{`effective film thickness' at point of closest approach}}$.

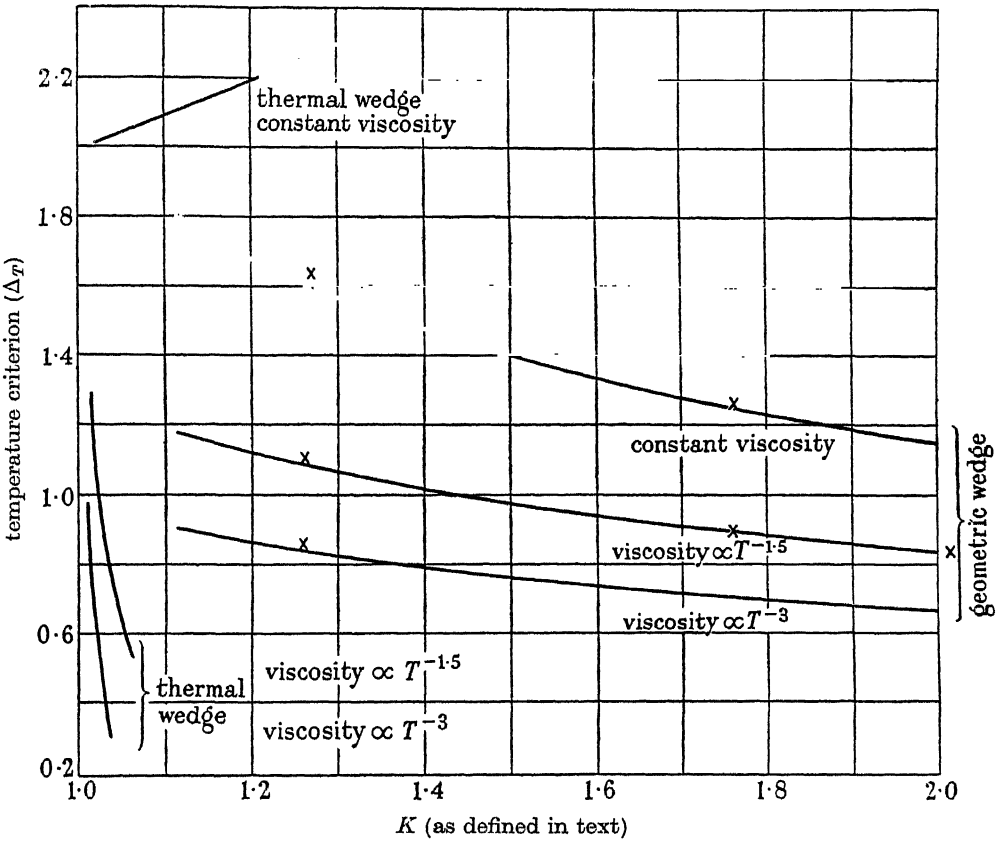


FIGURE 3. × = points calculated with allowance for oil expansion, $\alpha = 10^{-3}$.

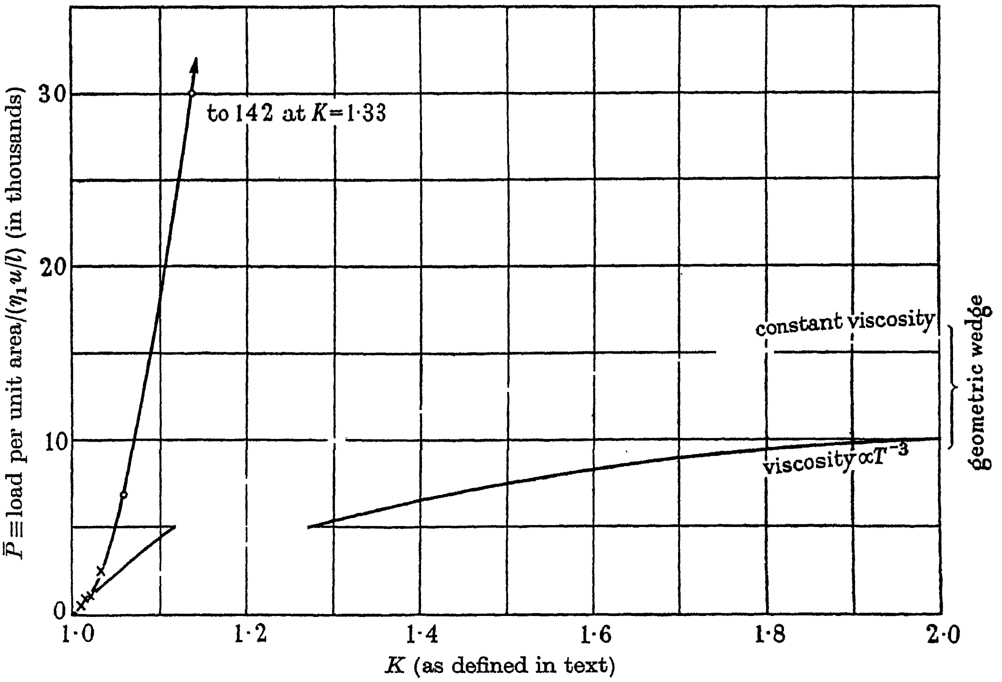
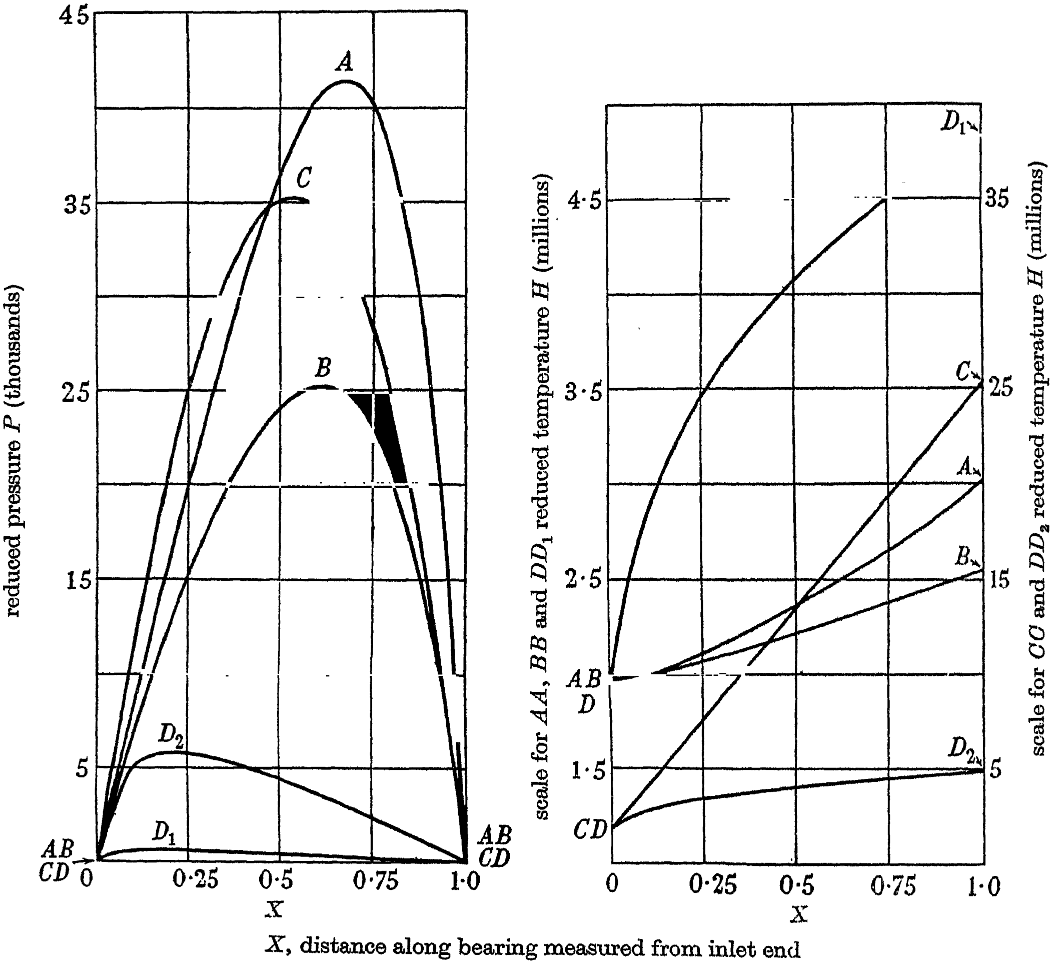


FIGURE 4. ○ = thermal wedge, constant viscosity. × = thermal wedge, viscosity ∝ T^{-3}.

Thus, when both geometrical and thermal wedges are present,

$$\begin{aligned}K &= \frac{H_1}{H_2}+\frac{1}{R_2}-1\\ &= \frac{H_1}{H_2}+\frac{A(\Theta_2-\Theta_1)}{1-A(\Theta_2-\Theta_1)}\\ &= \frac{H_1}{H_2}+\frac{A}{H_2^2}\Delta_T\Big/\left(1-\frac{A}{H_2^2}\Delta_T\right).\end{aligned} \tag{11}$$



a *b*

FIGURE 5. *a*, curve DD_2D is curve DD_1D to 10 times the scale. *b*, geometric wedge $K = 2$, constant density. Curve AA, constant viscosity; curve BB, viscosity $\propto T^{-3}$. Thermal wedge. Curve CC, constant viscosity, $K = 1{\cdot}33$; curve DD, viscosity $\propto T^{-3}$, $K = 1{\cdot}03$. Curve DD_2 is curve DD_1 plotted to right-hand scale.

From this definition the value of K has been calculated. The solutions of equations (10) under the conditions stated above are given in table 4 and plotted in figures 1 to 5.

The load, drag and friction criteria have been plotted against K in figures 1, 2 and 3 respectively. The quantity $\overline{P}$, defined by Δ_L/H_2^2 or by

$$(\text{load per unit area})/(\eta_1 U/l),$$

has been plotted against K in figure 4. And P and Θ have been plotted against X in figure 5 for values of the several variables representing a heavily loaded bearing.

TABLE 4

K	h_2/l	α	β	Δ_L	Δ_D	Δ_μ	Δ_T
10/9	10^{-3}	0	0	47×10^{-3}	0·95	20	1·81
10/9	10^{-3}	0	−1·5	32	0·63	20	1·18
10/9	10^{-3}	0	−3	24×10^{-3}	0·48	20	0·91
1·25	10^{-3}	0	0	88×10^{-3}	0·90	10	1·63
1·27	10^{-3}	10^{-3}	0	95	0·90	9·5	1·65
1·25	10^{-3}	0	−1·5	61	0·61	10	1·09
1·26	10^{-3}	10^{-3}	−1·5	64	0·61	9·6	1·10
1·25	10^{-3}	0	−3	48	0·48	10	0·84
1·26	10^{-3}	10^{-3}	−3	49×10^{-3}	0·48	9·6	0·85
1·5	10^{-3}	0	0	131×10^{-3}	0·84	6·4	1·41
1·5	10^{-3}	0	−1·5	95	0·60	6·3	0·98
1·5	10^{-3}	0	−3	75×10^{-3}	0·47	6·3	0·77
1·75	10^{-3}	0	0	151×10^{-3}	0·80	5·3	1·26
1·76	10^{-3}	0	0	154	0·80	5·2	1·27
1·75	10^{-3}	0	−1·5	112	0·58	5·2	0·89
1·76	10^{-3}	10^{-3}	−1·5	115	0·58	5·1	0·90
1·75	10^{-3}	0	−3	92×10^{-3}	0·47	5·1	0·71
2	10^{-3}	0	0	159×10^{-3}	0·77	4·9	1·16
2	10^{-3}	0	−1·5	122	0·57	4·7	0·84
2·01	10^{-3}	10^{-3}	−1·5	123	0·57	4·7	0·84
2	10^{-3}	0	−3	101	0·47	4·6	0·67
1·02	10^{-3}	10^{-3}	0	11×10^{-3}	1·00	92	2·02
1·01	10^{-3}	10^{-3}	−1·5	4	0·64	147	1·29
1·01	10^{-3}	10^{-3}	−3	3×10^{-3}	0·48	196	0·97
1·06	$6{\cdot}32\times10^{-4}$	10^{-3}	0	28×10^{-3}	1·00	36	2·06
1·02	$\sqrt{40}\times10^{-4}$	10^{-3}	−1·5	6	0·47	79	0·95
1·02	$\sqrt{40}\times10^{-4}$	10^{-3}	−3	3×10^{-3}	0·32	120	0·64
1·13	$4{\cdot}47\times10^{-4}$	10^{-3}	0	60×10^{-3}	1·00	16·7	2·13
1·04	$\sqrt{20}\times10^{-4}$	10^{-3}	−1·5	7	0·35	52	0·72
1·02	$\sqrt{20}\times10^{-4}$	10^{-3}	−3	3×10^{-3}	0·22	88	0·44
1·33	$3{\cdot}16\times10^{-4}$	10^{-3}	0	142×10^{-3}	1·00	7·0	2·35
1·06	$\sqrt{10}\times10^{-4}$	10^{-3}	−1·5	7	0·26	36	0·53
1·03	$\sqrt{10}\times10^{-4}$	10^{-3}	−3	2×10^{-3}	0·15	67	0·30

α is the coefficient of cubical expansion of the oil with temperature

$$\rho = \rho_1\{(1+\alpha(T-T_1)\}.$$

β is the index of T in a power law variation of viscosity with temperature

$$\eta = \eta_1\left(\frac{T}{T_1}\right)^\beta.$$

K is the ratio of inlet to minimum wedge thickness $K = \dfrac{h_1}{h_2}+\dfrac{B\Delta_T}{1-B\Delta_T}$,

where for shortness B is written for $\dfrac{\alpha\eta_1 Ul}{\rho_1 Ch_2^2}$.

COMPARISON OF GEOMETRIC AND THERMAL WEDGES

Even a cursory inspection of any of the first three figures will reveal the profound effect that change of viscosity has on the performance of the thermal wedge. For instance (figure 1), Δ_L is almost constant, with a value of about 1/400, for the lubricant with the more rapid rate of viscosity variation. It is, of course, obvious that an effect of this kind must occur, since a rise in temperature implies a decrease is viscous dissipation, this in turn implies that the total temperature difference is less than it would have been with a smaller change in viscosity, and therefore the density change, which produces the thermal wedge, is less; but its magnitude came as rather a suprise.

It would, however, be wrong to infer that this low value of Δ_L means that load-carrying capacity of a given thermal wedge bearing with a given lubricant is necessarily less than that of a geometric wedge of the same size. The mechanism of altering K is quite different; broadly speaking, the thermal wedge increases K by decreasing H_2, whereas the geometric wedge tilts its pads and H_2 remains relatively constant. It follows that the load-carrying capacity of the former will be roughly inversely proportional to H_2^2, of the latter roughly independent of H_2. The relative load-carrying capacity is therefore more nearly exhibited by figure 4, since $\overline{P}$ is a direct measure of this quantity for a fixed size of bearing. This figure shows that even with a large variation of viscosity the two types of wedge are about equal on a load-carrying basis, and that with a small change of viscosity the thermal wedge completely outclasses the geometric.

The question of whether in actual practice the close approach envisaged can actually be achieved is rather outside the scope of this paper. The absence of tilting arrangements implies a more robust and mechanically simpler construction, which lends itself better to the production of plane surfaces. On the other hand, the distances envisaged (about 10^{-3} cm.) are such that really efficient filtration is needed.

The first three figures also show that the performance of a geometric wedge at medium and large values of K is not appreciably improved by the addition of the thermal effect. This, of course, explains why the existence of the thermal wedge has gone unsuspected for so long, since the calculations show clearly that it is only under conditions of really close approach that the density variation can contribute substantially to K.

The difference in pressure and temperature distribution between the two types of wedge under heavy load is delineated in figure 5 both for constant viscosity (curves AA and CC) and for a large variation of viscosity (curves BB and DD). The pressure curves show both that the point of maximum pressure is farther forward, and that its movement forward is greater for the thermal wedge. The temperature curves give an idea of the acccuracy of the assumption, made in several approximate treatments, that Θ is a linear function of X. It is seen that the assumption is only accurate for the thermal wedge if the viscosity be constant. As the load decreases it becomes more nearly true for both types of bearing.

CONCLUSIONS

Successful film lubrication requires a 'wedging action' of some kind. This can be achieved in two ways: (1) geometrically, by decreasing the width of the film in the direction of motion, and/or (2) thermally, by decreasing the density of the fluid in the direction of motion. The first (the geometric wedge) is more important than the second in many applications, but the second (the thermal wedge) is of value and may even be important if certain conditions are fulfilled. The conditions are (1) small variation of viscosity with temperature, which is desirable on other grounds; (2) large coefficient of cubical expansion for the lubricant; and (3) small film thickness, which implies good workmanship.

If these conditions can be simultaneously fulfilled then the thermal wedge may altogether outclass the geometric.

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES

Bisacre, F. F. P. & Bisacre, G. H. 1947 *Proc. Instn Mech. Engrs, Lond.*, **157**, 452.
Bower, G. S. Contribution to Fogg, A (1946).
Cameron, A. & Wood, W. L. 1946 Parallel surface thrust bearing. *Proc. 6th Int. Congr. Appl. Mech.*
Christopherson, D. G. 1941 *Proc. Instn Mech. Engrs, Lond.*, **146**, 126.
Cope, W. F. 1942 *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1903.
Fogg, A. 1946 *Proc. Instn Mech. Engrs, Lond.*, **155**, 49.
Harrison, W. J. 1912–23 *Trans. Camb. Phil. Soc.* **22**, 6.
Kingsbury, A. 1897 *J. Amer. Soc. Nav. Engrs*, **9**, 267.
Lamb, H. 1924 *Hydrodynamics*, p. 551. Cambridge University Press.
Reynolds, O. 1886 *Phil. Trans*, A, **177**, 157.
Shaw, M. C. 1947 *Trans. Amer. Soc. Mech. Engrs*, **69**, 381.
Swift, H. W. 1932 *Proc. Instn Civ. Engrs*, **233**, 267.
Goldstein, S. (Editor) 1938 *Modern developments in fluid dynamics*, **1**, ch. IV. Oxford University Press.

Crystal growth from solution

By S. P. F. Humphreys-Owen *
Imperial College, London

(*Communicated by Sir George Thomson, F.R.S.—Received* 13 *September* 1948—
Revised 18 *November* 1948—*Read* 17 *February* 1949)

[Plates 5 and 6]

Single crystals of the cubic form of sodium chlorate are studied in growth from aqueous solution. An interferometric technique measures, via refractive index, concentrations and gradients of concentration contiguous to the crystal. The rates of advance of individual faces are measured and compared with predictions derived from the Nernst supposition, namely, that the concentration is saturation at a crystal face and that the facial rate of advance depends only on diffusion. It is found that faces sometimes grow at rates consistent with this supposition, but that on other occasions smaller rates of growth are observed, including instances of complete stoppage of growth. It is also found that individual faces can change discontinuously from one rate of growth to another. The observation of Berg, that concentration and normal concentration gradient are functions of position at the face, is developed, and empirical equations are obtained for the functions which connect the rate of growth with the concentration and gradient distribution along the face. To account for the advancing surface remaining plane, notwithstanding the non-uniformity of normal concentration gradient along it, it is necessary to suppose a transport of material along the face. Some observations on dissolution have been made, and a non-uniformity of concentration along a dissolving face is found and discussed.

1. Introduction

A crystal growing in solution obtains its material by diffusion of the solute. Because the diffusion process is slow in comparison with crystallization, material removed by the growing crystal is not instantaneously replaced, and there results an impoverishment of the solution near the crystal and a system of concentration gradients around it. It was assumed by Nernst that the activity of the crystal maintained the concentration at its surface at the saturation value, and that growth depended only on the presentation of material by diffusion. Early attempts to verify this assumption employed high rates of stirring with the object of reducing the zone of non-uniform concentration to a small residual layer. Fick's law of diffusion could then be employed across the residual layer without integration. But the results were not consistent, mainly because of the uncertainty in the thickness of the residual unstirred layer.

Valeton (1923, 1924) considered theoretically a crystal growing in an infinite, unstirred, solution. In order to account for the observed differences of rate of growth amongst different types of face (indicating that diffusion is not the only operative factor), a suggestion due to Berthoud (1912) was employed. It was assumed that the concentration, c_a, at the crystal was intermediate between the value, c_∞, far from the crystal, and the saturation value, c_s, and that growth was proportional to this local supersaturation. The equation was used

$$G_a = k(c_a - c_s), \tag{1}$$

---

* Now at Birkbeck College, London.

where G_a is the mass per unit area per second entering the crystal, and k is a constant dependent only on the nature of the crystal surface and which can be different for different types of surface. For mathematical simplicity a polygonal crystal was assumed to be equivalent to a spherical crystal of radius a. The flow per second across any spherical surface in the solution is, from Fick's law, $4\pi r^2 D(dc/dr)$, where D is the diffusion constant of the solution and r is the radius of the surface measured from an origin at the centre of the crystal. This flow is, in the stationary state, independent of r, and the concentration at r is obtained by integrating between the boundaries $c = c_\infty$ at $r = \infty$ and $c = c_a$ at $r = a$, whence

$$c = c_\infty - (a/r)\,(c_\infty - c_a).$$

From Fick's law and the definition of G_a we have

$$G_a = D\frac{dc}{dr}_{r=a},$$

whence
$$G_a = (D/a)\,(c_\infty - c_a). \tag{2}$$

The radial rate of growth is da/dt, and $G_a = \rho\, da/dt$, where ρ is the density of the solid crystal. The state of affairs is represented by simultaneous solution of (1) and (2), whence

$$\frac{da}{dt} = \frac{(D/\rho)\,(c_\infty - c_s)}{a + D/k}, \tag{3}$$

$$c_a = \frac{c_\infty D/k + ac_s}{a + D/k}. \tag{4}$$

Equations (3) and (4) show how simultaneous measurement of rate of growth and concentration can test the comparative validity of the Nernst and Berthoud assumptions. The Nernst assumption is that k is very large compared with D, so that c_a tends to c_s, and (3) becomes the purely diffusive

$$a\frac{da}{dt} = (D/\rho)\,(c_\infty - c_s). \tag{3a}$$

The first simultaneous measurement of rate of growth and concentration was made by C. W. Bunn (1932, unpublished work at I.C.I. Ltd.). The crystal used was cubic sodium chlorate, and it was found that the problem was more complex than expected because the concentration was not uniform along a crystal face; it had a minimum at the face centre. It was also found that faces of the same type in one crystal could have different rates of growth, indicating that growth can depend on something in addition to the type of face and the diffusive environment. Berg (1938), in continuation of Bunn's work, noted also that the component of the concentration gradient normal to the face (which is proportional to the exit of solute per unit area per second from the solution) was not uniform along the face, being a maximum at the face centre. Since the face grows as a plane, there must be a transport of material along the face in a layer too small to be observed. He suggested that this layer might be identified with a Volmer adsorbed phase (Volmer & Estermann 1921).

The present paper describes further work on sodium chlorate crystals, employing the technique of Bunn and Berg.

2. EXPERIMENTAL TECHNIQUE

Figure 1 shows, not to scale, the arrangement of the apparatus and the physical dimensions of certain parts. A small crystal of side about 0·03 cm. and height about 0·01 cm. is grown in a film of supersaturated solution between two partially reflecting optical plates. If the film is wedge-shaped, Fizeau fringes are formed in the film. Crystal and fringes are photographed in plan through a microscope, using low magnification of from 10 to 20 linear.

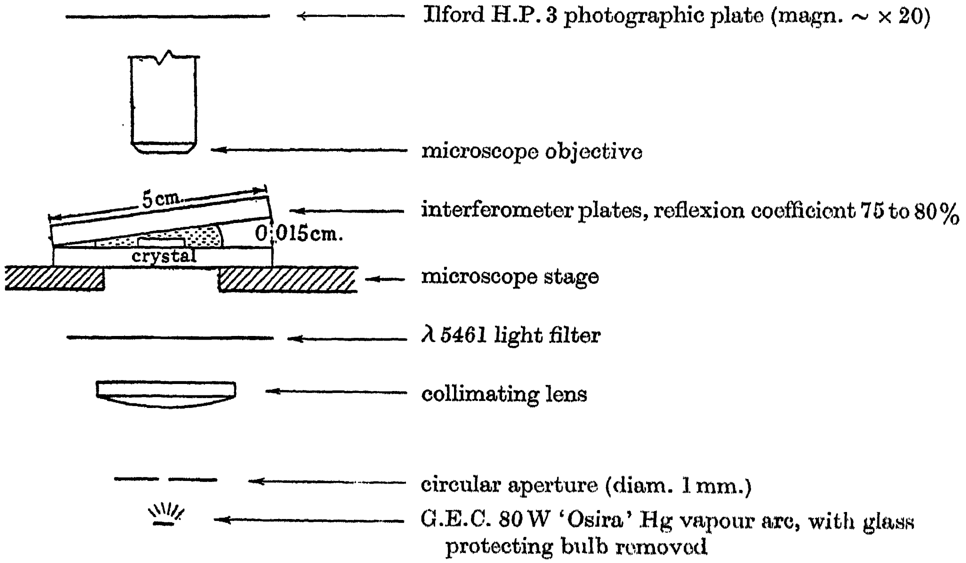
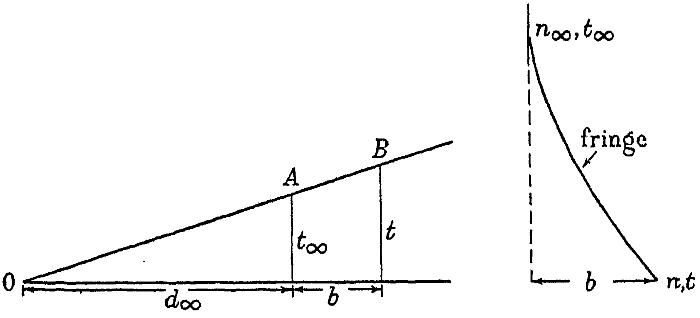

Ilford H.P. 3 photographic plate (magn. ~ × 20)
microscope objective
5 cm.
0·015 cm.
crystal
interferometer plates, reflexion coefficient 75 to 80 %
microscope stage
λ 5461 light filter
collimating lens
circular aperture (diam. 1 mm.)
G.E.C. 80 W 'Osira' Hg vapour arc, with glass protecting bulb removed

FIGURE 1. Diagrammatic arrangement of the apparatus, not to scale.

The concentration gradients around the crystal distort the fringes and are measured thereby. In addition, the rates of advance of the four free faces are measured, and the desired simultaneous observation of concentration and rate of growth is thus achieved. Let refractive index be n, and height of the wedge t (see figures 2 *a* and *b*).



a. Elevation *b*. Plan

FIGURE 2. Diagrams to illustrate the relation between deviation of a fringe and change of refractive index.

Consider a fringe which, far from the crystal, is formed along a straight line of constant wedge height t_∞, at a constant distance d_∞ from the apex of the wedge. As the crystal is approached, and the refractive index is reduced from its value, n_∞, far from the crystal to a less value, n, near the crystal, the fringe, which is a locus of constant optical path nt between the plates, is deviated in the direction of increasing wedge height a distance b, say, and is formed where the wedge height is t. We have

$$n_\infty t_\infty = nt,$$

therefore

$$n_\infty d_\infty = n(d_\infty + b),$$

whence

$$n = n_\infty \frac{d_\infty}{d_\infty + b}. \tag{5}$$

Equation (5) was used to measure the refractive index at any point on any fringe. n_∞ is for the solution initially used and is measured separately. d_∞ is measured by means of the travel of the microscope stage, high accuracy not being necessary. b is the important measurement and is measured as accurately as possible on the photographic plate. The magnification of the system must be known, taking care to observe any aberration over the field of view. The optical plates must be of a planeness such that fringe distortion caused by departure from planeness is less than the error in locating a fringe. The planeness must hold good over the whole plate surface because (5) assumes that BA in figure 2a passes through the apex O, which will not be the case if there is any curvature of the surfaces.

A variation of the arrangement, in which the plates were mutually parallel, was also used. Fringes are then formed which are loci of constant refractive index, and are located where $2nt$ is an integral number of wave-lengths, t being now constant. The difference of n between two successive fringes is $\lambda/2t$, where λ is wave-length, so that if N fringes are counted from the region of uniform n_∞ far from the crystal to a fringe near the crystal where the refractive index is n, we have

$$n_\infty - n = N\lambda/2t.$$

In practice the order of interference far from the crystal will not necessarily be integral, and we must write

$$n_\infty - n = (N + e)\,\lambda/2t, \tag{6}$$

where e is a fraction which can be estimated from the tint of the region of uniform refractive index.

High accuracy requires close fringe spacing, and hence, with given gradients, large t. But the fringes degrade in contrast and eventually vanish if t is increased too much. The utility of this method is therefore limited. But it affords the same accuracy as the wedge method when high rates of growth, providing large gradients, are employed. Since n is a function of the concentration, the fringes are contours of equal concentration, and the method is therefore convenient for qualitative interpretations of photographs.

Absolute error in these methods is about ± 2 in the fourth decimal place of n. Relative error in one photograph is less and can be kept down to ± 5 in the fifth decimal place.

The crystals in the present work were allowed to form spontaneously in the supersaturated film, glass spacers being used to maintain the separation between the plates. This method has the advantage that the crystal is undisturbed throughout its history, but there is a disadvantage in that the crystals do not always grow up to fill the space between the plates. Diffusion can then take place over the top of the crystal and affect the results. Check experiments were, however, made in which the upper plate rested directly on a pre-grown crystal, and it was established that the main features of the results were not affected by the presence of a small diffusion path over the top of the crystal.

3. Measurement of relationship between refractive index and concentration

Solutions were made up from sodium chlorate 'Pure Crystals' supplied by Hopkin and Williams. The solute was recrystallized once and solutions were filtered free of dust. n_∞ was measured on a Pulfrich refractometer made by Adam Hilger Ltd. Illumination in all the present work was the λ 5461 radiation from a mercury-vapour arc. The results were converted to a standard temperature of 18° C by employing the measurements on temperature variation of n for sodium chlorate aqueous solutions given by Meiers & Isaac (1906), and are presented in table 1.

Table 1. Refractive index of $NaClO_3$ aqueous solutions for λ 5461 light

room temperature (° C)	concentration (g. solute per 100 g. solution)	refractive index	refractive index at 18° C
18·7	36·30	1·37429	1·37432
19·0	42·54	1·38204	1·38225
27·4	45·98	1·38488	1·38687
27·4	46·97	1·38600	1·38799
26·3	47·99	1·38763	1·38944
27·4	48·97	1·38876	1·39075
27·4	50·12	1·39036	1·39235
27·4	50·83	1·39130	1·30329
27·4	51·99	1·39286	1·39485
26·8	52·99	1·39433	1·39620
19·3	55·99	1·40011	1·40027

Over the range of concentration 46 to 56 % the relationship is linear in accordance with the empirical equation

$$n_{18^\circ\,\mathrm{C}} = 1{\cdot}32414 + 0{\cdot}00136c \quad (46\,\% < c < 56\,\%),$$

but below 46 % the relationship departs from the above linearity.

Concentrations were expressed as percentage supersaturated $(c - c_s)$ by the use of figures for the solubility of $NaClO_3$ taken from *The International Critical Tables*, which yield a linear relationship with temperature in accordance with

$$c_s = 44{\cdot}35 + 0{\cdot}232t \quad (0^\circ\,\mathrm{C} < t < 40^\circ\,\mathrm{C}).$$

4. THE RATES OF GROWTH OF INDIVIDUAL FACES

The experimental results were compared with a simplified treatment based on Valeton's method (§ 1). The crystal of rectangular plan was assumed to be replaceable by a circular crystal of radius a. A uniform, constant concentration, c_∞, was assumed to exist at and beyond a radius r_0 measured from the centre of the crystal. This assumption will be later justified. In two dimensions, the flow in the stationary state across a circumference of radius r is $2\pi r D(dc/dr)$, and is independent of r. Thenceforward the treatment follows similar lines to that given in § 1 and we obtain

$$a\,da/dt = \frac{(D/\rho)\,(c_\infty - c_a)}{\log_e(r_0/a)}, \tag{7}$$

$$a\,da/dt = \frac{(D/\rho)\,(c_\infty - c_s)}{\log_e(r_0/a) + (1/a)\,(D/k)}, \tag{8}$$

$$c_a = \frac{(1/a)\,(D/k)\,c_\infty + c_s \log_e(r_0/a)}{\log_e(r_0/a) + (1/a)\,(D/k)}. \tag{9}$$

Strictly there would not be a stationary state (upon which assumption the above is based), unless there were a source at r_0 and a did not increase with time. But an approximately stationary state in practice exists, for it is observed experimentally that, except when the crystal is very small, the fringe system changes only slowly with time. The region surrounding the crystal, in which c is non-uniform, extends slowly at a rate of the same order of magnitude as the rate of growth of the crystal. It is found that r_0 maintains an approximately constant ratio of 3 to 1 to the half-side of the crystal. Thus, in place of attempting the rigorous solution of the true diffusion equation in $\partial c/\partial t$, actual rates of growth were investigated in the light of (7) and (8), employing a constant value for the term $\log_e(r_0/a)$.

Rates of growth of individual faces were measured by the advance of each face from the estimated nuclear point of the crystal (usually recognizable by a lack of transparency caused by irregular lattice formation in the first instants of growth). The distance, y, thus advanced by a face was made equivalent to the radius, a, of the circular crystal, where a was assumed to equal the mean of y and the half-diagonal $y\sqrt{2}$.

Rates of growth were first investigated in the light of the Nernst assumption, namely, that $D/k = 0$, and $c_a = c_s$. In this case (7) and (8) become

$$a\,da/dt = \frac{(D/\rho)\,(c_\infty - c_s)}{\log_e(r_0/a)}. \tag{10}$$

The ratio D/ρ has been measured by a method to be described later. Its value is $6{\cdot}8 \times 10^{-8}$ cm.²/sec. $a\,da/dt$ is equivalent to $\frac{1}{1{\cdot}39}\frac{d}{dt}y^2$. If the value of r_0/a is taken as 3, the logarithm term is 1·1. (10) is then equivalent to

$$\frac{d}{dt}y^2 = 8{\cdot}67 \times 10^{-8}(c_\infty - c_s). \tag{10a}$$

Equation (10a) suggests testing the constancy of $(d/dt)\,y^2$, which will be given the symbol L hereafter. It was, in fact, found that L was constant for individual faces, but that it often had different values amongst the faces of a crystal. It was also found that L could change from one value to another in the course of a face's history. Figure 3 is the record of growth over a period of 40 min. of the four free faces of one crystal. The constancy of L is seen, as well as its differing values amongst the faces.

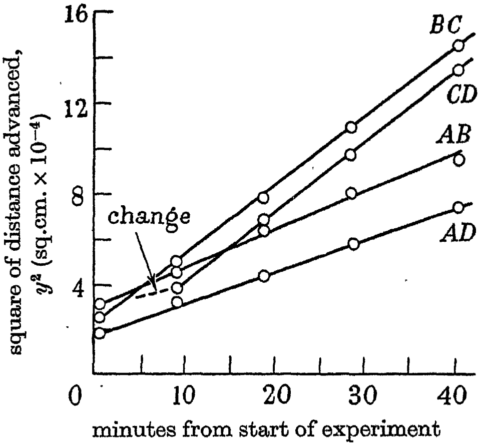


FIGURE 3. The rates of change of y^2, the square of the distance advanced from the crystal nucleus, for the four faces of an individual crystal. Face CD underwent a change of rate of growth between the first and second exposures.

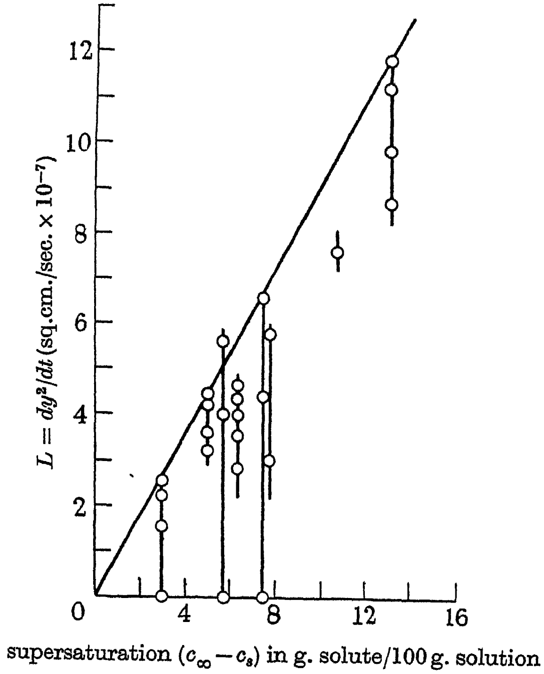


FIGURE 4. Relationship between $L = dy^2/dt$, and the supersaturation $(c_\infty - c_s)$ prevailing far from the crystal. Points for individual crystals are joined by vertical lines.

One change of L is shown. It is, of course, unexpected that all faces (when they have different rates of growth) should preserve constant L; if they were all obeying the Nernst assumption they should have equal rates of growth. This point will need to be raised later.

Equation (10a) was tested numerically by plotting L for a number of experiments against the main supersaturation $c_\infty - c_s$. The result is shown in figure 4. It is seen that, notwithstanding the variety of values of L for a given $c_\infty - c_s$ (including three cases of complete stoppage of growth), there is a maximum value of L for a given supersaturation, which is proportional to the supersaturation. This proportionality is shown by the straight line through the origin in figure 4. Points corresponding to individual crystals, i.e. to single values of the supersaturation, have been joined by vertical lines. (Some points are missing because of failure to measure L on account of too many changes.) The slope of the straight line through the origin is $8{\cdot}67 \times 10^{-8}$ cm.^{2}/sec., and thus it appears that faces corresponding to points on this line are growing in accordance with the Nernst assumption. In order to distinguish between these faces and those of less L, we may rewrite (10a)

$$L_{\text{max.}} = 8{\cdot}67 \times 10^{-8}(c_\infty - c_s). \tag{10a}$$

The other aspect of the problem is the concentration at the face, and this will be dealt with in the next section.

5. CONCENTRATION AT THE CENTRE OF A GROWING FACE

It was found by Bunn and Berg that the centre concentration, c_m, at the face was a minimum. c_m therefore reflects most directly the activity of the face and will be considered first.

c_m has been found in the present work to be saturation within experimental accuracy for those faces obeying (10a). Thus it appears that the mechanism of growth which causes a distribution of concentration along the face is not important in determining the rate of growth of the face; faces which obey (10a) grow quantitatively as if the concentration at them were uniform and had the saturation value actually observed at their centre.

Now faces which do not obey (10a), and have less L values, are found to have values of c_m greater than saturation. If it be generally assumed that faces behave as though they had a uniform concentration c_m along them, c_m can be substituted for c_a in (7), (8) and (9), and it can be investigated whether these equations are valid in general, as well as for the special case (10). (7), for example, does predict an inverse correlation between L and c_m for a given crystal, and, in terms of observed quantities, we test

$$L = 8{\cdot}67 \times 10^{-8}(c_\infty - c_m), \tag{11}$$

(11) being a general equation of which (10a) is a special case.

Figure 5 is a plot of L against $c_\infty - c_m$. Points corresponding to individual crystals have been given separate symbols and have been joined. It is seen that there is an inverse correlation between L and c_m, as predicted by (11), but that (11) is not accurately obeyed. If it were, all the points should lie on the same straight line through the origin, the line being identical with that of figure 4 (corresponding to (10a)). There are two other discrepancies, namely, that L and c_m should be time-dependent because of the term $1/a$ in the denominators of (8) and (9), whereas they are observed to be constant with time. (The question of the constancy of L has already been mentioned.)

It is considered, however, that all three discrepancies can be assigned to one cause, the loss of central symmetry when faces of one crystal have different rates of growth and different concentrations at them. The analogy of the circular crystal is a valid approximation only when all four faces behave similarly and the plan of the crystal remains a square. In general this is not the case, and three types of error are introduced. First, the fixed nuclear point from which the y's are measured loses its significance when faces have advanced different distances from it, and hence the values of L are in error. Secondly, the tendency of differently growing faces to have different concentrations at them causes a modification of these concentrations since all faces are connected by the same diffusion field. Thirdly, these modifications in themselves modify the rates of growth.

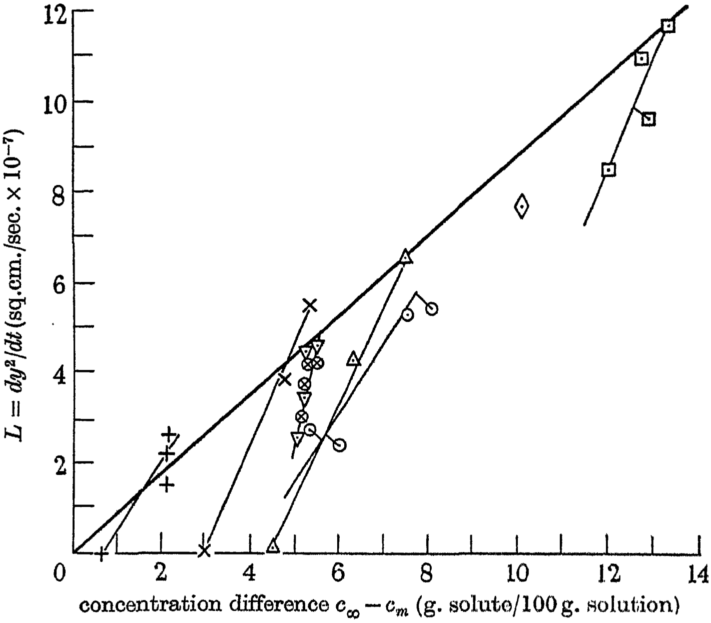


FIGURE 5. Relationship between $L = dy^2/dt$, and the concentration difference $(c_\infty - c_m)$ between the region far from the crystal and the face centre. Points for individual crystals are given separate symbols.

The basic assumption of all the above treatment, the equation (1), cannot therefore be accurately tested without modifying the experimental technique so as to isolate a single face from the others. This is possible but has not yet been done. But in the meanwhile it can be said that the results of the present work are not inconsistent with (1). They suggest that there is a 'free' mode of growth in which the Nernst assumption holds good and (10a) is obeyed, but that often a face is in some way hindered, and has a lower value of the constant k in (1). It has been mentioned that a face can change its L value in the course of its history. It is significant that this is observed to happen abruptly. The fringe system at the face is seen to rearrange itself in a few seconds. This point will be discussed in the concluding section.

Some of the effects which have been mentioned are illustrated in the group of six photographs of one growing crystal in plate 5. The interferometer plates were parallel, and the fringes outside the crystal are contours of equal concentration. The outline of two faces in (*e*) has been roughly traced with a pin. Otherwise the photographs are untouched and the rectangular outline is that of the crystal. Fringes within this outline are caused by irregularities on the crystal's upper surface. These were caused by a burst of upward growth between the first and second exposures. There was also another crystal growing in the neighbourhood, whose surrounding fringe system has gradually obtruded into the field of view. These defects would normally cause the rejection of the experiment, but many effects are visible in this one set of photographs, and it has accordingly been selected for reproduction.

The concentration decreases as the crystal is approached, and the fact that the concentration is higher at the corners than at the face centres is seen. The left-hand face (as mounted in the prints) underwent a decrease of rate of growth at a moment after the first exposure, and then reverted to its original behaviour after the second exposure. The change of concentration at this face is vividly seen. The lower face grew more slowly than the others throughout, and in the third exposure, was photographed during a period of complete stoppage. In this photograph the absence of concentration gradient normal to the face is seen.

6. DISTRIBUTION OF CONCENTRATION AND NORMAL CONCENTRATION GRADIENT ALONG GROWING FACES

A growing face appears as a straight line in the focal plane of the apparatus. This line will be taken as the axis of a position co-ordinate x, and the origin will be taken at the centre of the line. It was found by Berg that g, the component of the concentration gradient normal to the face, was a function of x. There must therefore be a flow of solute along the x-axis which converts the intake from outside, which is a function of x, into an intake into the solid crystal which on the time average is independent of x. Otherwise the face would not grow as a plane. Let (see figure 6) there be a flow $F = F(x)$ g./sec. in the x-direction in a layer of thickness h, and width t (the interferometer gap).

Let the rate of advance of the face be R cm./sec. in the direction normal to its length. Into an element dx of the layer there enters from outside a mass $tDg\,dx$ g./sec. Into the solid crystal below there enters a mass $t\rho R\,dx$ g./sec. The net flow out of the element via the layer is $(dF/dx)\,dx$ g./sec. From conservation of matter we have

$$dF/dx = t(Dg-\rho R). \tag{12}$$

The processes at the face may be cyclic, as successive crystal planes are laid down, but the fact that a stationary fringe system is observed at the face indicates that these processes can be replaced by an average steady state represented by (12). Let the half-length of the face be l, then, equating the total intake per second from outside to the total mass per second incorporated into the solid face, we have

$$tD\int_{-l}^{+l} g\,dx = t\rho R\int_{-l}^{+l} dx.$$

Writing $\int_{-l}^{+l} g\,dx \Big/ \int_{-l}^{+l} dx = \bar{g}$, where $\bar{g}$ is the mean normal gradient, we obtain

$$D\bar{g} = \rho R, \tag{13}$$

whence $$dF/dx = tD(g-\bar{g}),$$

and $$F = tD\int_0^x (g-\bar{g})\,dx. \tag{14}$$

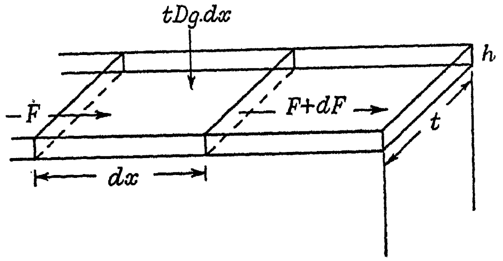


FIGURE 6. Diagram to illustrate the relation between the lateral flow, F, along a face and the component of concentration gradient normal to the face, g.

The layer flow per unit area (the area taken normal to the direction of flow) is

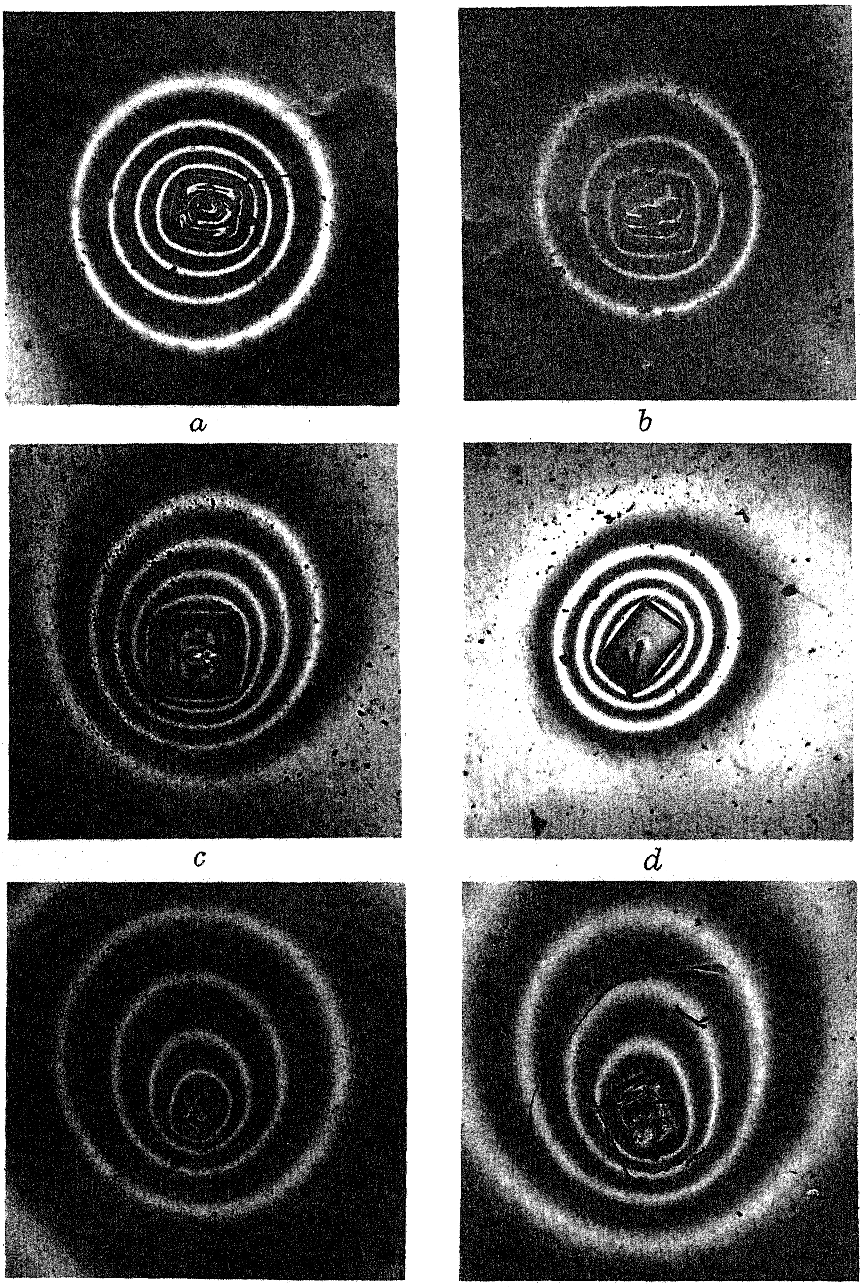
$$F' = (D/h)\int_0^x (g-\bar{g})\,dx. \tag{15}$$

Figure 7 is a sketch of the type of curve $g(x)$ which is obtained and the resulting curve $F(x)$. The layer flow is directed from the centre towards the corners, changes direction (is algebraically zero) at the centre, is a maximum where $g = \bar{g}$, and, on the assumption of no flow round corners, is zero at the corners. Its value at some point P is represented graphically by the shaded area in figure 7, the area being taken as negative below the line $g = \bar{g}$.

From (13), a graph of R against $\bar{g}$ is a straight line of slope D/ρ. Such a graph for one crystal's history is shown in figure 8. This was the method employed to measure D/ρ, mentioned in § 4.

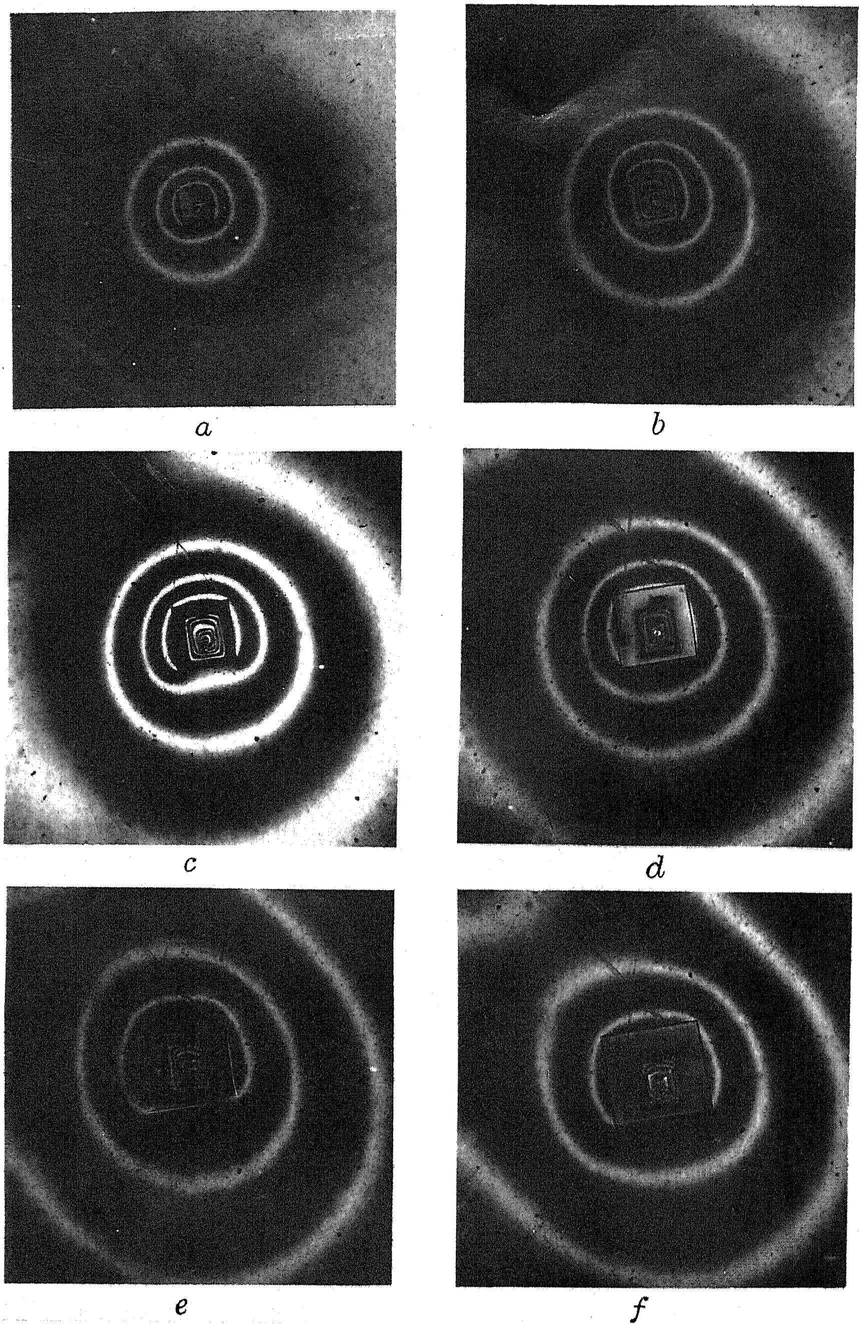
In figure 9, g has been plotted against distance along the four faces of individual crystals, and the faces extended into the same straight line. The lack of symmetry about the face centre is caused by the existence of the common diffusion field, enhanced by some diffusion over the top of the crystal. (When adjacent faces possess mutually unequal gradient distributions, some distortion of the $g(x)$ curve is thus caused.)

Undistorted $g(x)$ curves are found to fit the equation to a parabola. This equation, although possibly only an approximation to the true theoretical relationship, is



a *b* *c* *d* *e* *f*

Six stages in the growth of one crystal. The rectangular outline is that of the crystal. Appearances within it are caused by surface irregularities. The outline of two faces in *e* has been roughly traced, otherwise the photographs are untouched. Outside the crystal, the fringes are contours of equal concentration. Note the change of concentration at the left-hand face in *b*, and the stoppage of growth of the lower face in *c*.



Six different dissolving crystals. The blunted rectangular outline is that of the crystal. The photographs are untouched. Appearances within the outline are caused by surface irregularities. Outside the outline the fringes are contours of equal concentration. Note the lower concentration at the corners except when, as in the lower portions of *e* and *f*, dissolution has proceeded far enough to round off the faces.

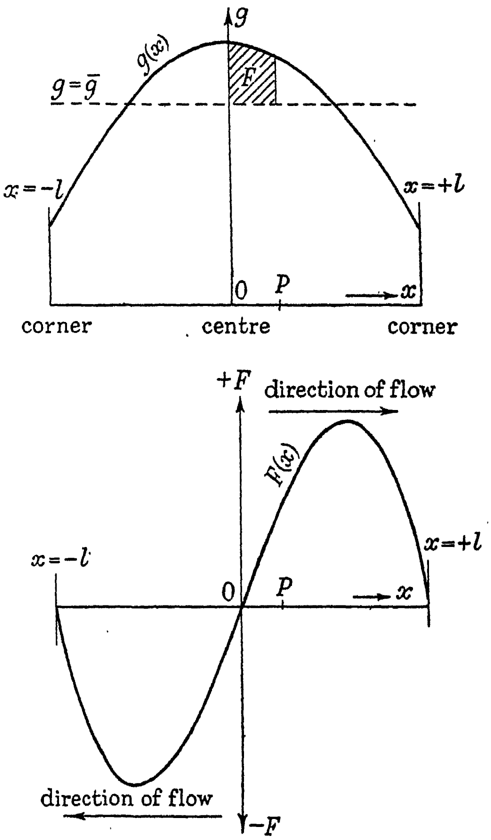


FIGURE 7. Relation between the form of the normal gradient function, $g(x)$, and the form of the layer flow function, $F(x)$, along a face.

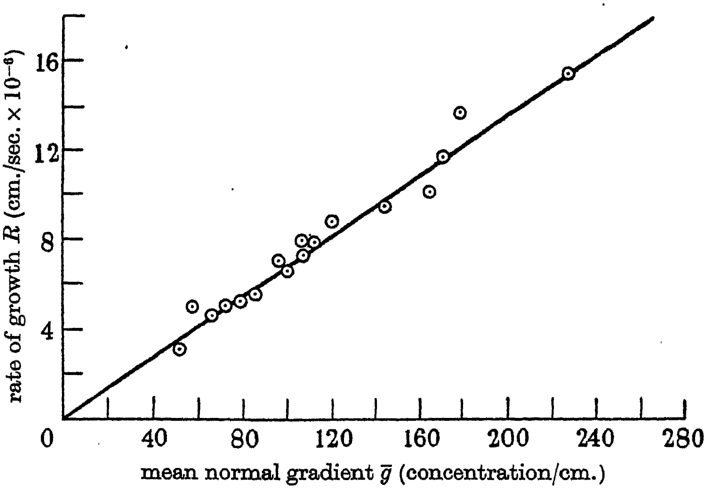


FIGURE 8. Relation between rate of growth and mean normal gradient of concentration at a face.

convenient as an empirical representation. Let the normal gradient at the face centre be g_m, and that at the corners be g_e. Then the parabola is

$$g = g_m - (g_m - g_e)\,(x^2/l^2).$$

From (13),
$$\bar{g} = g_m - \frac{g_m - g_e}{3} = \rho R/D.$$

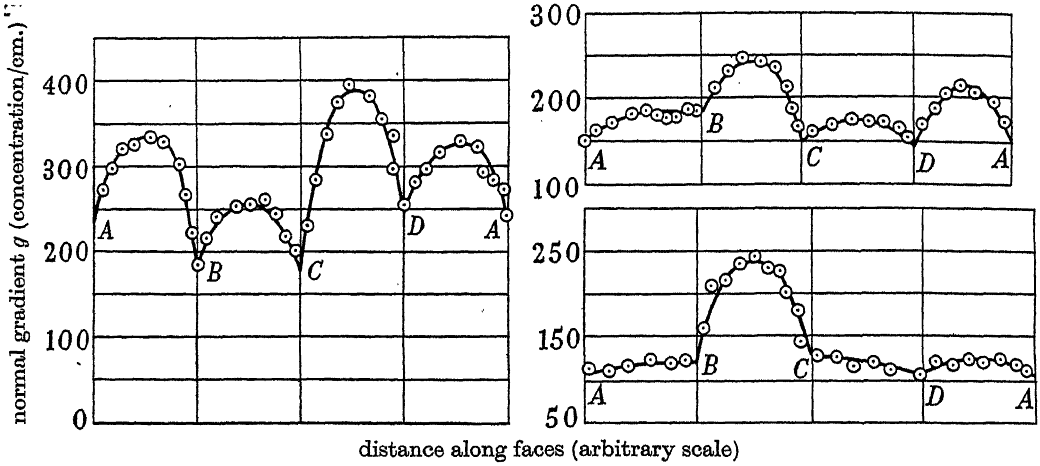


FIGURE 9. Relation between normal gradient of concentration and position along faces. The three graphs are from different experiments and have no correlation with each other.

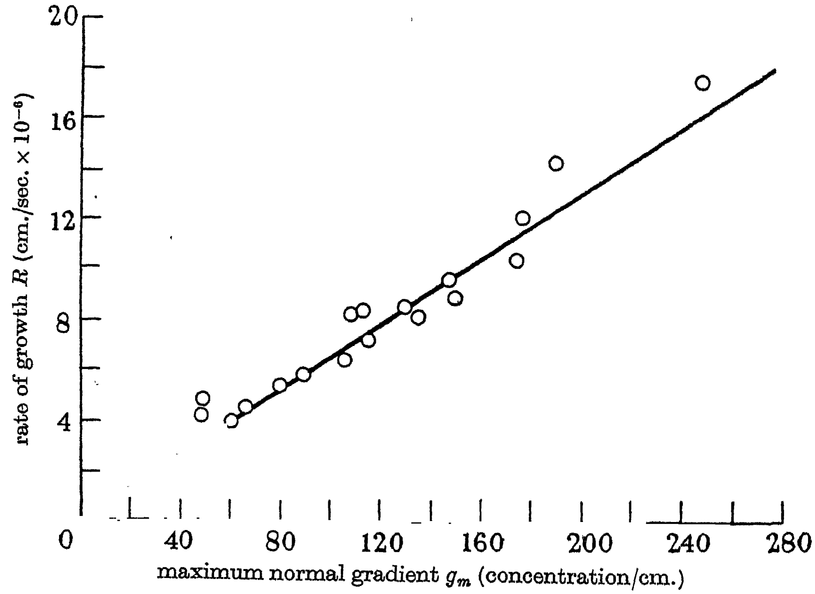


FIGURE 10. Relation between normal gradient at face centre, g_m, and rate of growth, R.

Another experimental relation is required to fix the parabola in terms of observed quantities, and is provided by the fact that g_m is proportional to R. Figure 10 illustrates this proportionality. Let $g_m = qR$. Then the parabola can be written

$$g = R\left\{\frac{\rho}{D} + \left(q - \frac{\rho}{D}\right)\left(1 - 3\frac{x^2}{l^2}\right)\right\}. \qquad (16)$$

From (14)
$$F = tR\left\{(Dq-\rho)\left(x-\frac{x^3}{l^2}\right)\right\}. \tag{17}$$

The maximum layer flow is where $g = \bar{g}$, i.e. where $x = l/\sqrt{3}$, and

$$F_{\text{max.}} = tR(Dq-\rho)\left(\frac{2l}{3\sqrt{3}}\right).$$

Putting in the numerical values

$$\begin{aligned} t &= 0{\cdot}01\,\text{cm.}, & q &= 1{\cdot}62\times 10^7\,\text{sec./cm.}^2, \\ R &= 10^{-5}\,\text{cm./sec.}, & \rho &= 2{\cdot}5\,\text{g./cm.}^3, \\ D &= 1{\cdot}7\times 10^{-7}\,\text{g./cm./sec.}, & l &= 0{\cdot}015\,\text{cm.}, \end{aligned}$$

the order of magnitude of the layer flow becomes

$$1{\cdot}4\times 10^{-10}\,\text{g./sec.}$$

Now at the face centre DqR g./cm.^{2}/sec. enter from outside, whereas only ρR g./cm.^{2}/sec. crystallize locally. Thus about 9 % of the intake from outside at the face centre does not crystallize locally, but leaves via the layer.

For a comparison between layer flow per unit area and diffusive inflow per unit area, an estimate of the layer thickness h is required. At $g = \bar{g}$ the layer flow is a maximum, and its value per unit area is

$$(R/h)\,(Dq-\rho)\,(2l/3\sqrt{3}) = \frac{1{\cdot}4\times 10^{-8}}{h}\,\text{g./cm.}^2\text{/sec.}$$

The diffusive inflow at $g = \bar{g}$ is, per unit area,

$$\rho R = 2{\cdot}5\times 10^{-5}\,\text{g./cm.}^2\text{/sec.}$$

If h were 6×10^{-4} cm., the two kinds of flow per unit area would be equal. If h were of molecular dimensions, corresponding to a Volmer adsorbed layer, say 10^{-7} cm., the layer flow would be of the order 10^{-1} g./cm.^{2}/sec., about 6000 times greater than diffusive flow in the external solution. But since, with the present technique, refraction effects prevent observation much closer to the crystal than 10^{-4} cm., the layer flow could be of any magnitude between the above extremes without observable effects on the fringes. The question of the thickness of the layer, and of the flow per unit area in it, must therefore be left open.

The function $c(x)$, that of the concentration, will now be described. In this case also the symmetry about the face centre is destroyed when adjacent faces are behaving differently from each other. The five diagrams of figure 11 show contours of equal concentration drawn at intervals of 0·15 % concentration per contour. They are selected to illustrate cases of extreme differences of behaviour amongst faces (b, c, d and e), as compared with a, in which the faces are behaving similarly. Figure 11d shows two stopped faces, and e shows one. The difference between c_m and c_e (the corner concentration) is greater for higher L values. When L is zero (stopped faces), the difference $c_e - c_m$ is almost zero, but not quite, because of the linkage of adjacent faces via the common diffusion field. Figures 12a and b show the supersaturation at the face plotted as a function of x for two different crystals, with the four faces extended into the same straight line. In figure 12b, c_∞ is high, and the resulting high value of L is associated with a high value of $c_e - c_m$.

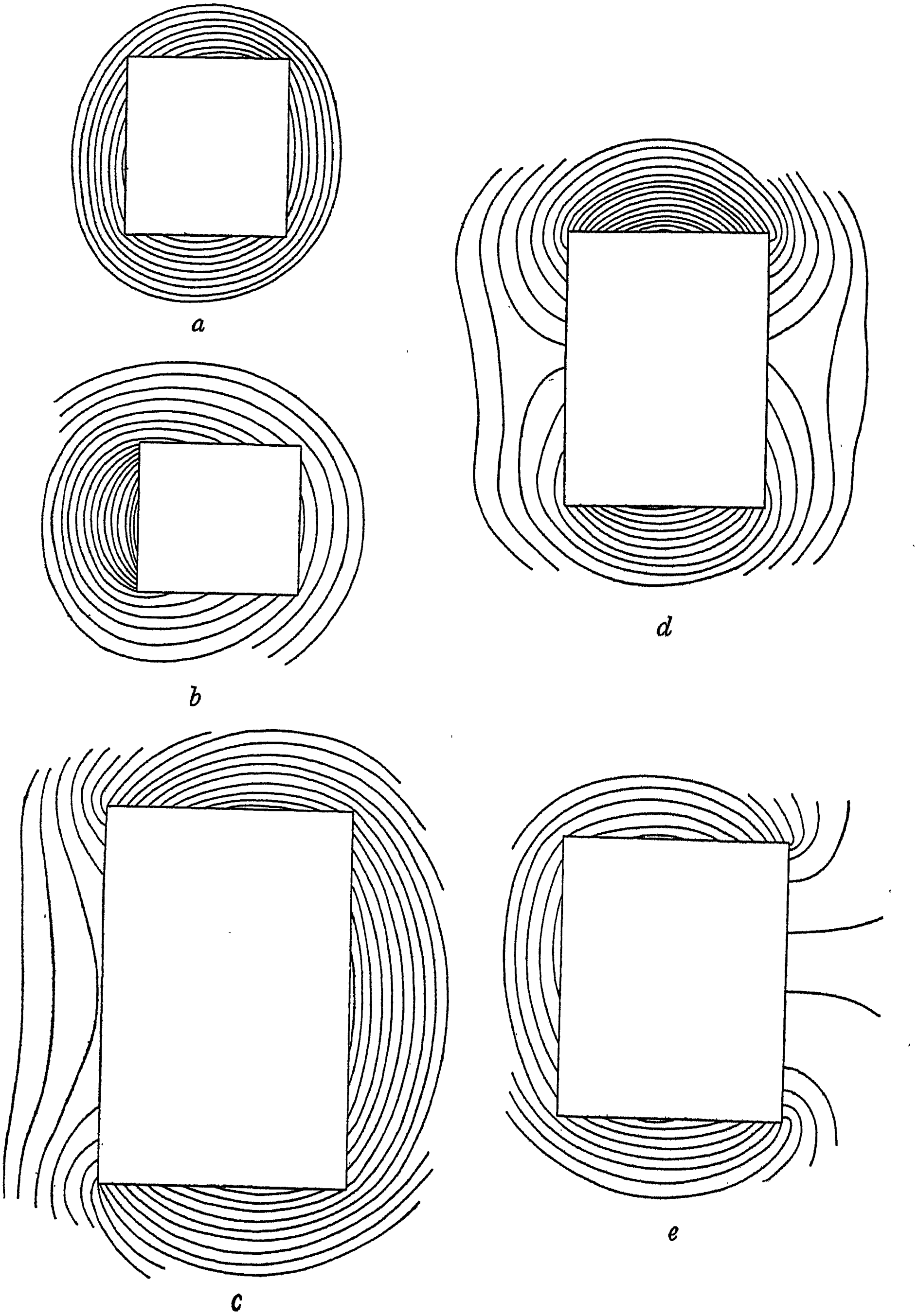


Figure 11. Contours of equal concentration round five different crystals, contrasting a, where differences in rate of growth between the faces is small, with the others, where large differences exist. Stopped faces are shown in d and e.

The difference $c_e - c_m$ was found to be proportional to L, as can be seen from the graph of figure 13, in which faces with $c(x)$ curves as far as possible undistorted have been selected. The proportionality is such that

$$c_e - c_m = 2{\cdot}1 \times 10^6 L. \tag{18}$$

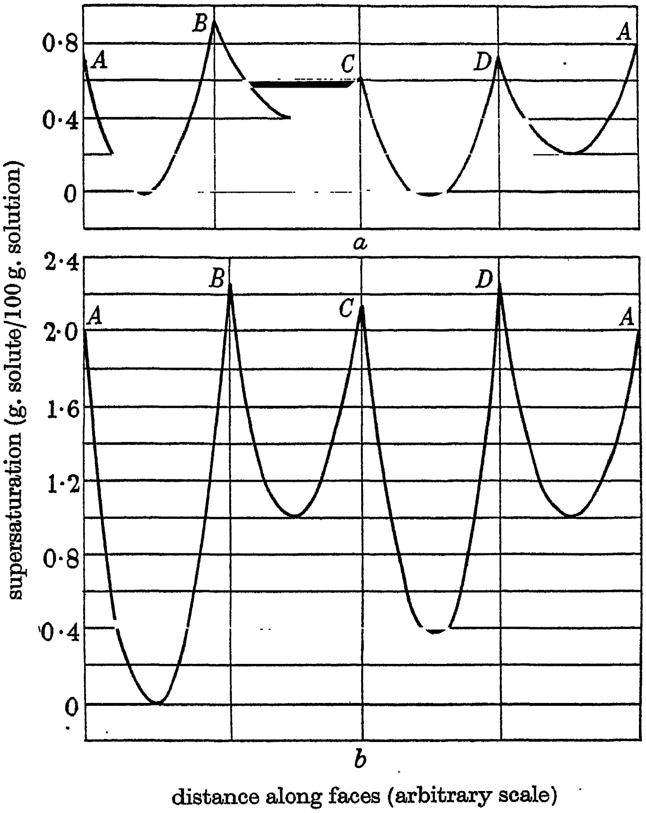


FIGURE 12. Relation between supersaturation at the face and position along the face. *b* is for an experiment in which the concentration far from the crystal was higher than in the case of *a*.

An empirical equation for $c(x)$ is conveniently obtained by making use of an observation reported by Berg and confirmed in the present work; the absolute gradient, i.e. the gradient normal to lines of equal concentration, is independent of x. Since at the face centre the absolute gradient equals g_m (the inflow being normal to the face), we have

$$\sqrt{[(\partial c/\partial x)^2 + g^2]} = \text{const.} = g_m.$$

Abbreviate (16) to $$g = \alpha - \beta(x^2/l^2),$$

where $$\alpha = g_m = qR \quad \text{and} \quad \beta = 3R(q - \rho/D),$$

then $$dc/dx = x\sqrt{\left(\frac{2\alpha\beta}{l^2} - \frac{\beta^2}{l^4}x^2\right)}.$$

Integrating and putting $c = c_m$ at $x = 0$, we obtain

$$c = c_m + l\frac{2\alpha}{3}\sqrt{\frac{2\alpha}{\beta}}\left\{1-\left(1-\frac{\beta}{2\alpha}\frac{x^2}{l^2}\right)^{\frac{3}{2}}\right\}. \tag{19}$$

Putting $l = y$, which is true when all faces are behaving similarly and is nearly true for the symmetrical $c(x)$ functions which were selected in obtaining (18), we have

$$2lR = 2y\frac{dy}{dt} = L.$$

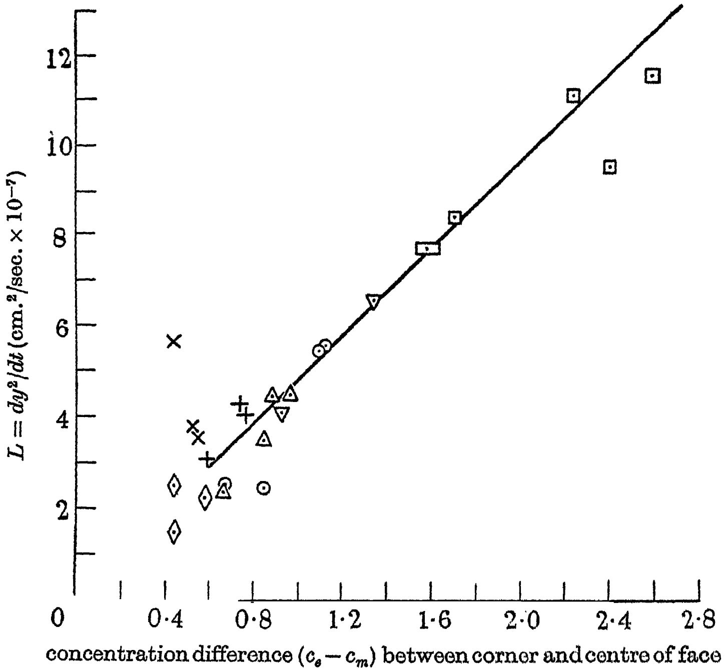


FIGURE 13. Relationship between $L = dy^2/dt$, and the concentration difference $(c_e - c_m)$ between the corner and the face centre. Points corresponding to individual crystals are given separate symbols.

Substituting numerical values for α and β, (19) becomes

$$c = c_m + 1{\cdot}45 \times 10^7 L\{1-(1-0{\cdot}139x^2/l^2)^{\frac{3}{2}}\}, \tag{20}$$

and with $x = l$

$$c_e - c_m = 2{\cdot}9 \times 10^6 L.$$

This agrees well enough with (18), and in fact (20) was found to fit the $c(x)$ curves well.

Since L, upon which $c(x)$ depends, is affected both by the external diffusion head, $c_\infty - c_s$, and by the 'hindrance' (a term introduced to denote the departure of a face from the Nernst behaviour), it will be of interest to separate the two effects on $c(x)$. Assuming the truth of (11) we have, from (10a),

$$L_{\text{max.}} = \mu(c_\infty - c_s),$$

from (11),

$$L = \mu(c_\infty - c_m), \quad \text{where} \quad \mu = 8{\cdot}67 \times 10^{-8}.$$

Writing $$H = L_{\text{max.}} - L = \mu(c_m - c_s),$$

and abbreviating (20) to $$c = c_m + Lf(x),$$

(20) can be rewritten $$c = c_s + L_{\text{max}}\, f(x) + H\{(1/\mu) - f(x)\}. \tag{21}$$

The second term in the right-hand side is the effect of the external diffusion environment, and the third term that of the 'hindrance'.

7. A NOTE ON DISSOLUTION

Plate 6 consists of six photographs of dissolving crystals taken with the optical plates parallel. The blunted rectangular outline is that of the crystal in each photograph. Appearances within the outline are caused by etching of the upper crystal surface by the undersaturated solution in which the crystals were placed.

The corners dissolve fastest, and the original rectangular outline becomes blunted and eventually circular. At the commencement of dissolution, before the rounding off of the corners has proceeded far, it is seen that the concentration is lower at the corners than elsewhere along the face. (In dissolution, of course, the concentration rises as the crystal is approached.) This non-uniformity of concentration is less marked than with growth and vanishes when the original outline has become circular, as is seen in the lower portions of *e* and *f*.

The crystals begin by dissolving evenly all round, but very soon the dissolution becomes asymmetrical. Cases of symmetrical dissolution are *a* and *b*, of moderate asymmetry are *c* and *d* and of marked asymmetry are *e* and *f*.

A full investigation of these effects would require a correlation of the rate of dissolution with the normal gradient everywhere at the surface, which would be more difficult than with growth because of the curvature of the surface. But in the meanwhile some remarks can be made.

If it is assumed that diffusion is not such a slow process relative to dissolution as it is relative to crystallization, it will be possible for diffusion to carry away material released by the dissolving crystal at a rate sufficient to counter the tendency of the crystal to bring the concentration up to saturation. This disposal of released material will be more effective at the corners where the angle through which diffusion can take place is greater. The concentration will then be lower at the corners because the disposal of material overcomes the greater rate of release there. When the face has become rounded both effects vanish; the corners are no longer distinct as regards rate of dissolution, and the greater angle for disposal is no longer there.

The observations support these views. It is found that the concentration at the face is less than saturation at the beginning of the experiments. It rises to saturation first along the centre portion of the face, and later, after the rounding off of the surface, it becomes saturation all along the face. That is, the facility with which diffusion can keep the concentration below saturation at the crystal decreases as the released material floods the neighbourhood.

For a given angle of disposal, the concentration at a faster dissolving portion of the crystal should be higher than at a slower dissolving portion. This is found to be the case when accurate measurements are made by the wedge technique and diffusion over the upper crystal surface is prevented. (This effect should not be sought in plate 6 because dissolution of the upper crystal surface masks it.)

The remaining observation, the asymmetry of dissolution, finds no explanation in the above. It is a persisting effect; once one corner has started dissolution faster than the others, that portion of the surface retains its lead in dissolution until the end.

One would expect that the increasing difficulty of removal of material by diffusion would eventually equalize rates of dissolution at all portions of the crystal surface. It may be that there is evidence here of fast dissolving portions of the surface releasing material into a mobile layer as well as into the contiguous solution. In this way the released material would be disposed of and also hinder dissolution elsewhere.

8. DISCUSSION

The 'Nernst' mode of growth depends only on presentation of material by diffusion. The observation that some cubic faces grow in this way raises the question: how would more unstable types of face, such as (110), succeed in growing any faster? We are not concerned here with the initial creation of the characteristic shape in the submicroscopic state (when energy considerations pertaining to the surface are more important than diffusion), but with artificially produced unstable faces which are observed to eliminate themselves by rapid advance. This could be a matter of experiment with the present technique; for example, it would be interesting to observe the regrowth of a crystal which had previously been dissolved into a circular shape.

Regarding the effect which has been termed 'hindrance', its two salient properties are its abrupt onset and changes, and its considerable duration. The first suggests a change on the surface, but is not consistent with the second, since such a surface change would not be expected to persist during the laying down of many new crystal planes. It might be caused by adsorption of foreign material into the mobile layer, but 'hindrance' is so much the general case that it is difficult to see how any faces could be free from it for an appreciable period. Experiments could be made on this technique with increasingly careful purification, as well as with the deliberate introduction of impurities.

Complete stoppage of growth seems to be an effect of a different order, and could be ascribed to adsorption of foreign material on the preferred nucleation point or points of the face (the corners, for example). This explanation is not satisfactory for 'hindrance', for the adsorbed matter would be incorporated or removed when fresh crystal planes are laid down. It is significant that when the restart of growth is observed after a period of stoppage, using the parallel interferometer plate method, there is seen a sudden fall of concentration at a single point, not at a corner. A fringe, representing the lower concentration, issues from that point on the face. This suggests, though it does not prove, that nucleation at the corners is the usual process, and is prevented during stoppage of growth. There is then a finite, but small, chance of nucleation elsewhere, which, when it happens, causes the incorporation of the corner-adsorbed matter and subsequent free growth.

Another observation suggesting corner nucleation is the upward growth of the horizontal face in the beginning of an experiment. With the plates parallel the interference tint in this face changes as the crystal grows upwards, and the change is in the form of bands sweeping across the surface from a corner. (This, again, is not conclusive; a crystal skew with respect to the plates would show the same effect if the growth were by direct incorporation of material evenly over the surface.)

The explanation of the $c(x)$ and $g(x)$ functions is probably to be sought, however, in some model in which 'steps' emerge from a nucleation point and interact with the mobile layer (Thomson 1948). But until some information can be obtained on where the preferred nucleation points are, and on the height and velocity of the propagated steps, it would be premature to attempt a theory.

Regarding the behaviour of crystals other than $NaClO_3$, it becomes difficult to employ the present technique when solutions cannot be held in a supersaturated state. $NaClO_3$ is particularly suitable in this respect. But the $c(x)$ function has been observed with one other crystal, potassium ferricyanide, in unpublished work at I.C.I. Ltd.

The author's thanks are due to Imperial Chemical Industries Ltd. for the loan of an internal report on crystal growth, and to Mr C. W. Bunn and Dr A. F. Wells of their research staff for helpful discussion.

REFERENCES

Berg, W. F. 1938 *Proc. Roy. Soc.* A, **164**, 79.
Berthoud, A. 1912 *J. Phys. Chem.* **10**, 625.
Meiers, H. A. & Isaac, F. 1906 *J. Chem. Soc.* **89**, 413.
Thomson, G. P. 1948 *Proc. Phys. Soc.* **61**, 403.
Valeton, J. J. P. 1923 *Z. Krystallogr.* **59**, 168, 335.
Valeton, J. J. P. 1924 *Z. Krystallogr.* **60**, 1.
Volmer, M. & Estermann, I. 1921 *Z. Phys.* **7**, 13.

Flexural vibrations of the walls of thin cylindrical shells having freely supported ends

By R. N. ARNOLD, D.Sc. AND G. B. WARBURTON
Department of Engineering, University of Edinburgh

(*Communicated by M. Born, F.R.S.—Received* 25 *October* 1948—
Revised 14 *December* 1948)

The paper deals with the general equations for the vibration of thin cylinders and a theoretical and experimental investigation is made of the type of vibration usually associated with bells. The cylinders are supported in such a manner that the ends remain circular without directional restraint being imposed. It is found that the complexity of the mode of vibration bears little relation to the natural frequency; for example, cylinders of very small thickness-diameter ratio, with length about equal to or less than the diameter, may have many of their higher frequencies associated with the simpler modes of vibration. The frequency equation which is derived by the energy method is based on strain relations given by Timoshenko. In this approach, displacement equations are evolved which are comparable to those of Love and Flügge, though differences are evident due to the strain expressions used by each author. Results are given for cylinders of various lengths, each with the same thickness-diameter ratio, and also for a very thin cylinder in which the simpler modes of vibration occur in the higher frequency range. It is shown that there are three possible natural frequencies for a particular nodal pattern, two of these normally occurring beyond the aural range.

INTRODUCTION

A thin cylindrical shell may vibrate in a variety of ways depending on the particular straining actions involved. The present paper is confined to that type of vibration in which bending and stretching of the shell predominate, as, for example, that associated with the tones emitted by church bells. Some of the vibration forms, which are possible under such conditions, are illustrated in figure 1. For sections perpendicular to the axis of the cylinder, the vibration consists of both radial and tangential movement, a number of stationary waves being formed around the circumference. In the simplest case ($n = 2$)* there are four positions at which the radial movement is zero. These will be referred to, for convenience, as circumferential nodes even although some tangential motion is known to exist. The form of vibration increases in complexity with the number of nodes, but theoretically there is no limit to the number which may be present. For a cylinder with free ends, the fundamental vibration form for a given circumferential arrangement exists when all cross-sections of the cylinder describe precisely the same motion. In such a case, each axial strip of the shell remains straight during vibration. Further complication, however, may result due to the superposition of waves in the axial direction. Some examples of this are given in figure 1*b* for the case in which the ends of the cylinder are maintained circular and no directional restraint is imposed, a condition which will be termed 'freely supported'.

---

* A list of symbols is given in appendix 1, p. 251.

It will be observed that certain cross-sections of the cylinder remain at rest, depending on the number of axial waves which are present. The positions of these define the axial nodes. Thus an infinite number of axial vibration forms are theoretically possible, each of which may be combined with a corresponding number of circumferential forms. To define any particular nodal arrangement, it is only necessary to specify the number of circumferential waves n and the number of axial half-waves m; figure 1*c*, for example, illustrates the vibration form $n = 3$, $m = 4$. It will be shown later that three distinct natural frequencies are associated with each nodal pattern, depending on the relative amplitudes of the three component vibrations executed by an element of the cylinder.

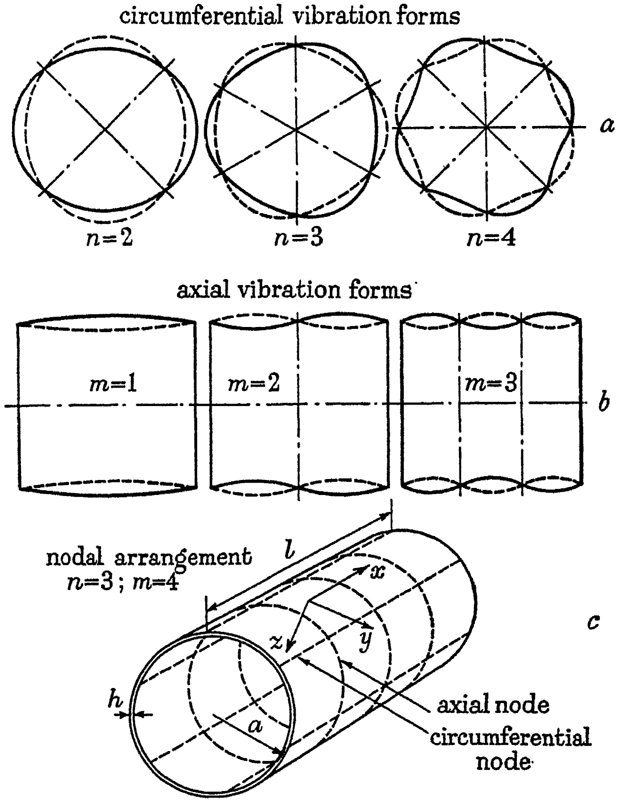


FIGURE 1. Forms of vibration of thin cylinders with freely supported ends.

THEORY*

Rayleigh (1894, chapter X) derived an expression for the natural frequencies of thin cylindrical shells with free ends vibrating with circumferential nodes. For this, he assumed the deformations to be inextensional; this means that the length of any line drawn on the middle surface of the shell (an imaginary surface situated at mid thickness) remains unaltered during vibrations of small amplitude. Bending was thus accepted as the predominant straining action, and stretching of the shell was ignored. While applicable to the particular problem which Rayleigh investigated, this concept has limitations. For example, while it may be legitimate to consider a line inscribed circumferentially on the middle surface to remain unextended during vibration, it is incompatible to assume in general that the middle surface as a whole is not subject to extension. Rayleigh did not discuss this aspect but merely confined himself to those vibrations in which all cross-sections throughout the entire length of the cylinder performed precisely the same motion. By equating kinetic energy at the mean position to strain energy at the maximum displacement, he obtained an expression for frequency in the form

---

* The mathematical analysis is given in the appendices.

$$f=\frac{1}{2\pi}\sqrt{\left\{\left[\frac{Eh^2}{12\rho a^4(1-\sigma^2)}\right]\left[\frac{n^2(n^2-1)^2}{n^2+1}\right]\right\}},$$

in which the second term expresses the nodal configuration.

For modes of vibration in which both circumferential and axial nodes exist, the assumption of inextension of the middle surface is inadmissible and the resulting analysis more complicated. In such cases, both bending and stretching of the shell wall must be considered.

The initial procedure involves the derivation of equations connecting the three component displacements u, v and w in the directions x, y and z (figure 1), of a point in the middle surface of the shell. This may be achieved in two ways:

(*a*) The equilibrium of a small element may be considered and the forces and moments acting on it expressed first in terms of strain and change of curvature of the middle surface, and finally as functions of the displacements. The resulting equations express u, v and w and their derivatives in terms of the elasticity and density of the material and the dimensions of the cylinder.

(*b*) A vibration form compatible with the end conditions may be assumed, the strain energy expressed in terms of the displacements and the kinetic energy in terms of the rates of change of the displacement. As both strain and kinetic energy are functions of the three independent variables u, v and w, the Lagrange dynamical equations may be used to form three equations connecting the displacements and their derivatives (see appendix 2).

It may be argued that the latter does not provide a direct attack, but it has the merit of brevity in the present case, when the form of the vibration is readily determined. Moreover, in the former method, a vibration form which will satisfy the derived equations must eventually be obtained. The equations resulting from either method will, however, only be as accurate as the original expressions adopted for the strain-displacement relations.

Procedure (*a*) has been followed by Love (1927), Flügge (1934) and Timoshenko (1940), but in view of the assumptions and approximations made by each during the analysis, the results are not identical. It appears advisable, therefore, before proceeding further to give a brief outline of the origin of the discrepancies. The three main differences appear to be as follows:

(i) Love and Flügge both make allowance, when considering an element, for the trapezoidal shape of the two faces, which lie in the planes perpendicular to the cylinder axis (figure 9, p. 252). Timoshenko assumes the inner and outer edges of these faces to be of equal length.

(ii) Each author makes use of different approximations in expressing the strains at a point distant z from the middle surface in terms of the extensional strain ϵ and change of curvature κ of that surface. Terms of the following types are obtained:

Love	Flügge	Timoshenko
ϵ	$\epsilon;\ \frac{z}{a}\epsilon;\ \frac{z^2}{a^2}\epsilon$	ϵ
$z\kappa;\ \frac{z^2}{a}\kappa$	$z\kappa;\ \frac{z^2}{a}\kappa$	$z\kappa$

In this, Love assumes that ϵ is small in comparison with $z\kappa$ and thus neglects terms of the type $\frac{z}{a}\epsilon$ and $\frac{z^2}{a^2}\epsilon$, but considers $\frac{z^2}{a}\kappa$. As the maximum value of z is $\frac{1}{2}h$ and the ratio of thickness to radius is small for a thin cylinder, terms $\frac{z^2}{a}\kappa$ and $\frac{z}{a}\epsilon$ are always small compared with $z\kappa$ and ϵ respectively. On this basis, Timoshenko elects to ignore all terms involving z/a.

(iii) Love considers the existence of a direct stress acting radially in the cylinder walls, satisfying the condition of zero stress at the inner and outer boundaries.

It seems relevant at this stage to indicate the objective of each of these analyses. That due to Timoshenko is only concerned with the straining action of thin cylinders, and terms involving inertia forces are absent. No attempt is therefore made to apply the result to problems of vibration. Love proceeds only so far as to indicate the type of functions which will satisfy the vibration equations, but does not investigate the implications arising from the use of such expressions. Only Flügge produces a direct calculation of frequency by obtaining the solution for a cylinder with freely supported ends. A strange fact, which emerges from his analysis, is the existence of three natural frequencies for any particular nodal pattern. Investigation reveals that each is associated with a unique arrangement of the ratios of the maximum vibration amplitudes occurring in the three component directions. While Flügge's equations allow for the presence of axial nodes, no attempt is made to investigate their existence.

Working independently the present authors derived expressions for frequency using the strain relations of Timoshenko in conjunction with the energy method described in paragraph (*b*), p. 240. They assumed that the axial, circumferential and radial vibration displacements were of the form

$$u = A\cos\frac{m\pi x}{l}\cos n\phi\cos\omega t,$$

$$v = B\sin\frac{m\pi x}{l}\sin n\phi\cos\omega t,$$

and
$$w = C\sin\frac{m\pi x}{l}\cos n\phi\cos\omega t,$$

which satisfies the conditions for freely supported ends and allows for the existence of axial nodes. It was later found that Flügge had made precisely similar assumptions.

In the above, A, B and C are constants, $\omega/2\pi$ is the frequency, and n and m are integers defining the circumferential and axial nodes; the number of nodes is respectively $2n$ and $(m+1)$ for freely supported ends.

From the three displacement equations, it is possible to eliminate A, B and C to form the cubic

$$\Delta^3 - K_2\Delta^2 + K_1\Delta - K_0 = 0,$$

where K_0, K_1 and K_2 are as specified in appendix 2, equation (15), and frequency

$$f = \frac{1}{2\pi a}\sqrt{\left\{\left[\frac{Eg}{\rho(1-\sigma^2)}\right]\Delta\right\}}.$$

It is found that the above equation supplies three real positive values for Δ, and consequently three frequencies are obtained for any given nodal configuration of the cylinder. The essential difference between these vibrations is the relative amplitudes of the motions u, v and w.

The above mathematical result does not offer any direct interpretation of the manner in which the frequency varies with the cylinder dimensions or nodal pattern. Direct calculation for specific cases was therefore undertaken. Four non-dimensional factors were chosen, namely,

$$\sqrt{\Delta} = \text{frequency} \times \sqrt{\left\{\frac{4\pi^2 a^2 \rho(1-\sigma^2)}{Eg}\right\}}.$$

$$\lambda = \frac{\text{mean circumference}}{\text{axial wave-length}} = \frac{2\pi a}{2l/m} = \frac{m\pi a}{l},$$

$$n = \text{number of circumferential waves},$$

$$\alpha = \frac{\text{thickness}}{\text{mean radius}} = \frac{h}{a}.$$

Of these, the first is proportional to frequency, the second to the number of axial waves and the fourth to the cross-sectional proportions.

Frequency calculations were performed for a series of thickness ratios α over a range of values of λ, and typical results are shown in figure 2*a* to *d*. In these, curves of $\sqrt{\Delta}$ are plotted to a base of λ for various nodal arrangements n, the value of α being constant for each diagram. For an infinitely thin cylinder (figure 2*a*) it will be seen that the natural frequency decreases as n increases, a result which one would not readily have predicted. That this frequency arrangement is not confined to hypothetical cylinders is shown by the remaining diagrams, where α has respectively the values 0·002, 0·01 and 0·0525. The last value, similar to that of a cylinder used in the experimental investigation, shows the same tendency at the higher values of λ, which are related to small axial wave-length. These theoretical results were so unexpected that a comprehensive experimental investigation was considered advisable.

EXPERIMENTAL INVESTIGATIONS

Steel cylinders were used for the experimental work in order that the natural frequencies could be stimulated by magnetic means. Difficulty was experienced in reproducing end conditions in accordance with the theoretical assumptions. The

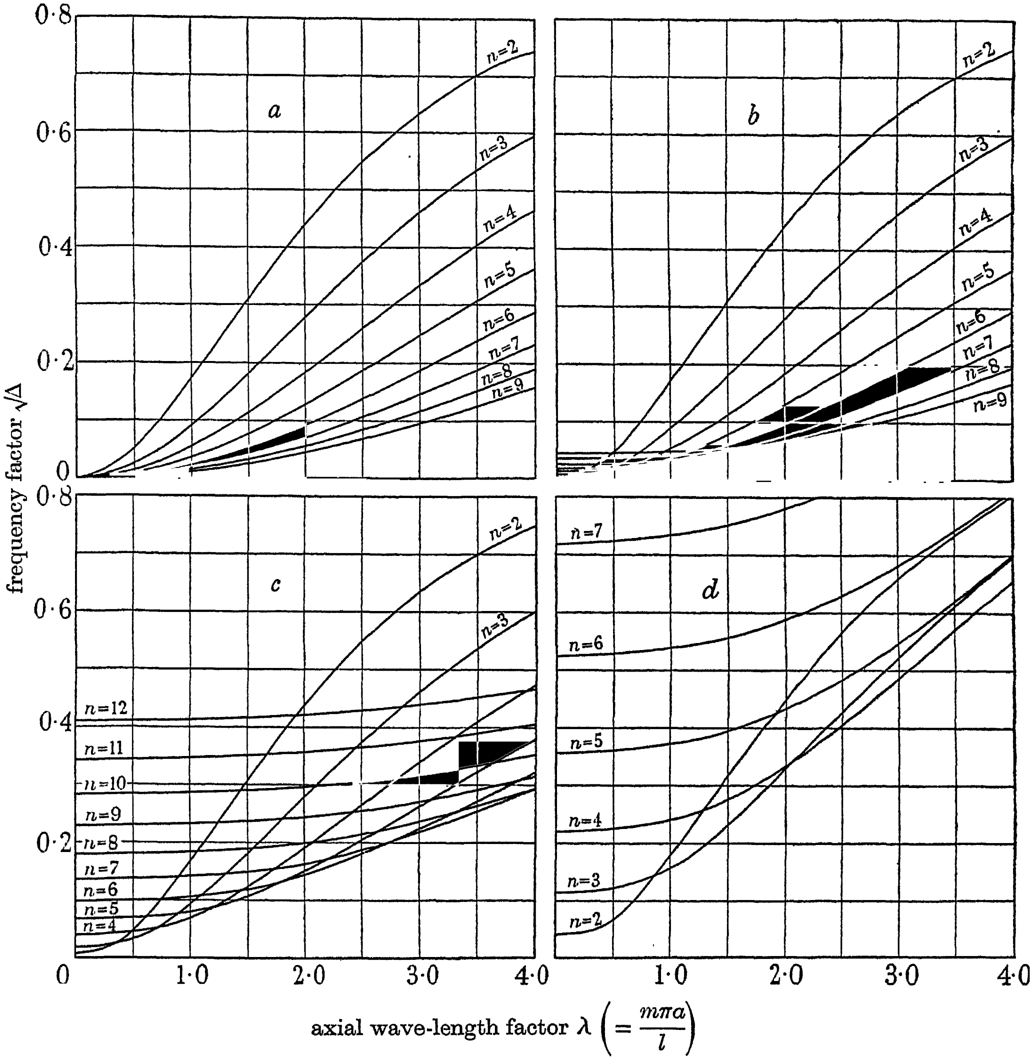


FIGURE 2. Theoretical frequency curves for cylinders with freely supported ends. α = thickness/mean radius. Values of α in a = 0, in b = 0·002, in c = 0·01 and in d = 0·0525.

type of end-piece finally adopted is illustrated in figure 3. It consisted of a circular steel plate accurately machined to fit the bore of the cylinder and shaped to provide as near as possible line contact with the inner surface. Slackness of fit in the end-pieces was found to produce inconsistent results. The cylinder, with end-pieces in place, was supported between centres so that it could be freely rotated and was set vibrating by a small electromagnet which was excited by an audio-frequency oscillator. The magnet was fixed to a support which could be moved to any axial position along the cylinder, and provision was made for adjusting the gap between magnet and cylinder.

When investigating a particular cylinder the magnet was placed midway along the axis and the frequency increased slowly by means of the oscillator. In passing through a natural frequency the cylinder was heard to respond and the oscillator was then adjusted in frequency so that the note emitted was of maximum intensity. On slowly turning the cylinder it was found that the sound intensity increased and decreased periodically, the number of fluctuations being a measure of the number of circumferential nodes. The nodes rotate with the cylinder due to the presence of minute imperfection, for if the cylinder were mathematically perfect, no preferential location of the circumferential nodes could exist. As a result of imperfection, two natural frequencies (Rayleigh 1894, p. 389) are present in association with two preferential nodal directions. For an accurately made cylinder, these frequencies are very close to one another, and this may lead to difficulty of interpretation if the resonance of each is not clearly defined. Only in a few cases could the results be determined by the unaided ear, and in general a medical stethoscope was used to listen to the response. For the axial nodes it was only necessary to move the stethoscope along the length of the cylinder and count the number of fluctuations of sound intensity. For doubtful responses, an electrical pick-up of condenser type was available; this recorded the vibration amplitude of the cylinder on a cathode-ray oscilloscope, but it was rarely necessary to make use of this instrument.

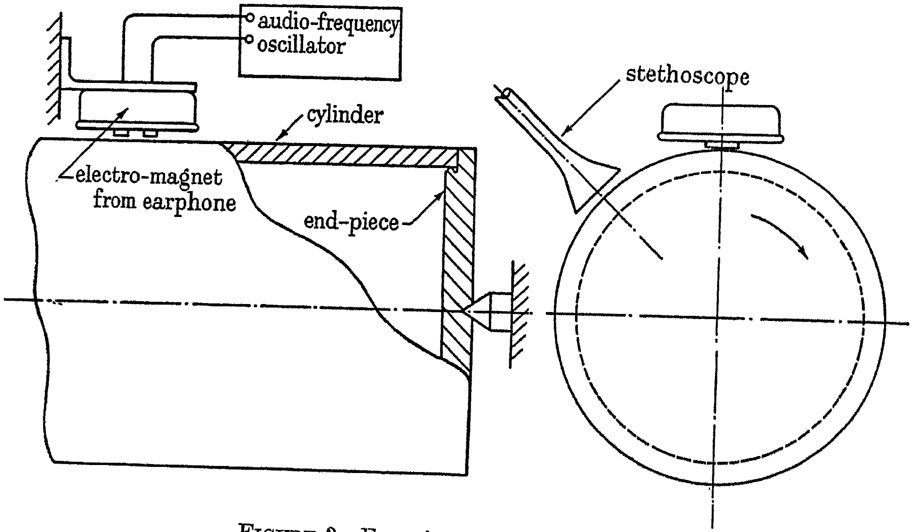


Figure 3. Experimental arrangement.

After progressing through the aural range of frequencies and recording the responses of the cylinder, the magnet was moved axially to another location and the procedure repeated to discover natural frequencies which had not been elicited at the mid-position. It was not possible for the magnet when situated centrally to elicit vibrations having a whole number of axial waves.

Some vibration configurations were found to be more readily stimulated than others, notably those having 6 or 8 circumferential nodes. The 4-nodal pattern at times presented difficulty, especially when its frequency was higher than those of the more complicated forms. It should be mentioned that the frequency values were slightly influenced by the axial pressure exerted on the supporting centres. Theoretically, a uniform axial compressive stress should not affect the frequency, but the pressure in this case resulted from the ends of the cylinder coming into contact with the end-pieces, and thus a small fixing moment was also introduced. This effect, however, was unimportant, as no appreciable error was introduced by the small axial forces used in the experiments.

Experimental work was conducted on five different cylinders. One of these, of thickness ratio $\alpha = 0{\cdot}01$, was specially designed to demonstrate the peculiar effects evident in figure 2*c*. The remaining cylinders ($\alpha = 0{\cdot}0525$) were identical in all dimensions except that of length. They were machined to give relative length ratios of 1, $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$ and thus allow a comparison to be made when each vibrated with the same axial wave-length.

EXPERIMENTAL RESULTS

The collected results for four experimental steel cylinders are plotted in figure 4. These cylinders were of outside diameter 3·950 in., internal diameter 3·748 in. and had the following effective lengths when freely supported at the ends:

cylinder	length (in.)
1	15·63
2	7·77
3	5·15
4	3·84

The full curves of figure 4 represent the calculated variation of frequency with λ for $n = 2$ (4 nodes) to $n = 6$ (12 nodes) and are similar to those of figure 2*d*. The calculations were based on the following constants: $E = 29{\cdot}6 \times 10^6$ lb./in.^{2}; $\sigma = 0{\cdot}29$; $\rho = 0{\cdot}283$ lb./in.^{3}. The experimental frequencies obtained from all the cylinders have been superposed on these curves, the points for cylinders 2 to 4 being numbered while those for cylinder 1 have been left unmarked. In the latter, it was possible to identify 41 natural frequencies, a list of which is given in table 1.

TABLE 1. EXPERIMENTAL FREQUENCIES (C./SEC.) CYLINDER NO. 1

number of axial half-waves	number of circumferential nodes ($2n$)				
	4	6	8	10	12
1	960	2,130	3,985	6,400	9,270
2	2,070	2,420	4,130	6,500	9,370
3	3,725	3,130	4,430	6,700	9,570
4	5,270	4,180	4,950	7,030	9,850
5	6,880	5,380	5,690	7,520	10,230
6	8,270	6,670	6,630	8,180	10,730
7	—	7,940	7,580	8,900	11,340
8	—	9,160	8,660	9,770	—
9	—	10,330	9,720	10,650	—
10	—	—	10,870	—	—

The experimental results are found to be in fair agreement with those predicted by theory, the greatest divergence being when $n = 2$. In this case, the experimental curve has been drawn dashed for comparison. These results are a striking demonstration of the strange phenomenon, to which reference has been made, namely, that certain vibration patterns may have frequencies higher than those related to vibrations which are much more complicated. It will be observed, for example, that at $\lambda \doteqdot 2{\cdot}3$ (wave-length 5·15 in.) the frequency for $n = 2$ is higher than those for $n = 3$, 4 or 5, and that the frequency for $n = 3$ has just overtaken that for $n = 4$.

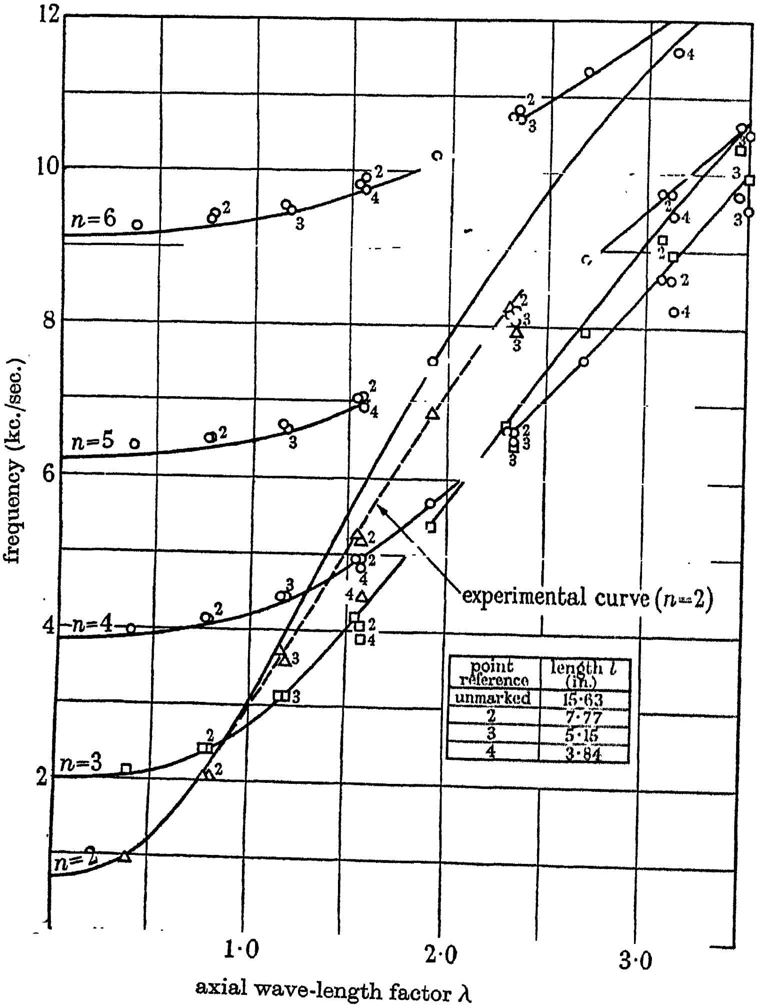


FIGURE 4. Natural frequencies of experimental cylinders ($\alpha = 0{\cdot}0525$; $h = 0{\cdot}101$ in., $a = 1{\cdot}925$ in.).

There is a tendency as λ is increased for each curve to continue to cross that immediately above it, though the reverse curvature evident on the curve for $n = 2$ shows that limitations must exist when λ is large.

The experimental results obtained from a thin cylinder 0·025 in. thick, 5·02 in. mean diameter and 2·065 in. long are shown in figure 5, where frequency is plotted to a base of circumferential nodes for both free and freely supported end-conditions. It was found possible in both cases to identify, within the aural range, all vibration patterns up to 32 nodes when no additional nodes were present in the axial direction.

For freely supported ends, it will be seen that the frequency initially decreases as the circumferential nodes increase, until a turning value is reached at approximately 14 nodes. Thereafter, the frequency increases with the number of nodes. The theoretical curve has been plotted for this condition, and while it is by no means coincident with that obtained experimentally, it is sufficiently close to corroborate the theory. It may be mentioned that the end-conditions were found to be extremely critical in this particular cylinder, the initial experiments being entirely unsuccessful owing to a small degree of slackness at the supports. It was only after special end-pieces had been accurately machined that the above results were obtained. The theoretical and experimental curves for the same cylinder, when the ends are free, are almost coincident. For such a case, the assumption of inextension is valid and the calculated results were obtained from the equation due to Rayleigh (1894, p. 386). In comparing the results for both end-conditions, it will be seen that beyond 20 nodes the frequencies are approximately the same. Below this, however, the curves rapidly diverge, giving, at 4 nodes, frequencies of 150 and 8350 c./sec. This is rather striking, for it shows that by the introduction of conditions which maintain circularity at the ends, the frequency is increased 55 times. This immense change is

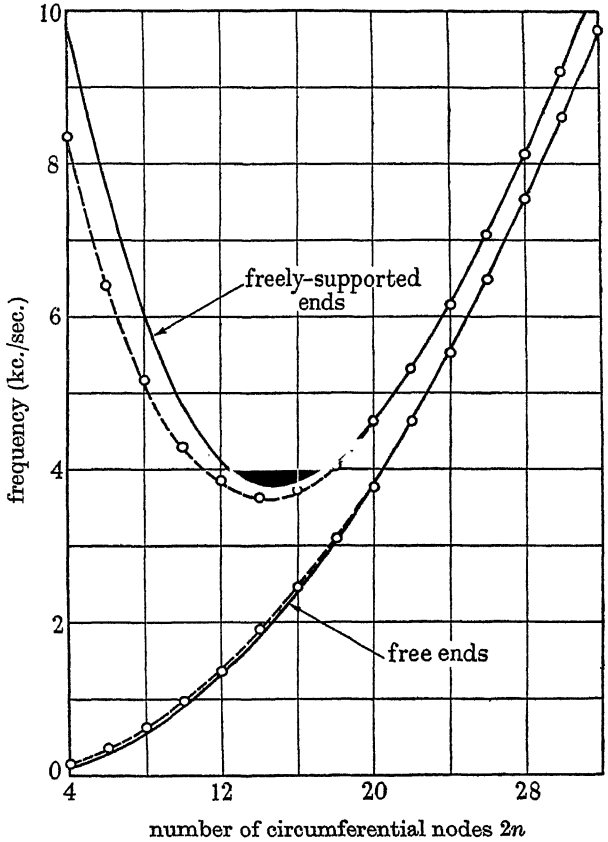


FIGURE 5. Experimental and theoretical frequency curves ($\alpha = 0{\cdot}01$).

a clear demonstration of the marked difference between extensional and inextensional vibration. It is due, as will be shown later, to the large strain energy involved when the cylinder walls are compelled to stretch.

Discussion

It has been shown that the higher frequencies may be associated with relatively simple modes of vibration, particularly in very thin cylinders, when the axial wave-length is small. In an attempt to understand the physical significance of this phenomenon, calculations were made to estimate the proportion of strain energy due to bending and stretching for various nodal arrangements. The calculations were based on equations (19) and (20) of appendix 3, and the results are shown plotted in figures 6 and 7. A comparison is made in figure 6 *a* and *b* between $n = 2$ and $n = 4$ over a range of λ for $\alpha = 0{\cdot}0525$, corresponding to the experimental cylinders for which frequencies are plotted in figure 4. It will be seen from figure 6*a* that with four circumferential nodes the strain energy is extremely small for values of λ up to 0·5, and that it is almost entirely composed of bending. At higher values, however, the stretching energy increases rapidly and becomes predominant, as may be observed from the shaded area representing the contribution of bending. With eight circumferential nodes (figure 6*b*), the situation is very different. In this case the bending effect is prominent throughout the whole range and never contributes less than 50 % of the total energy. Comparison shows that the total energy is approximately the same in each diagram when $\lambda = 1{\cdot}2$, even although the bending and stretching components are entirely different. This is in approximate agreement with figure 4, where it will be seen that the curves for $n = 2$ and $n = 4$ cross at $\lambda = 1{\cdot}3$. That these are not identical is due to the fact that the frequency is influenced by the kinetic energy function which is dependent on the component amplitudes A, B and C.

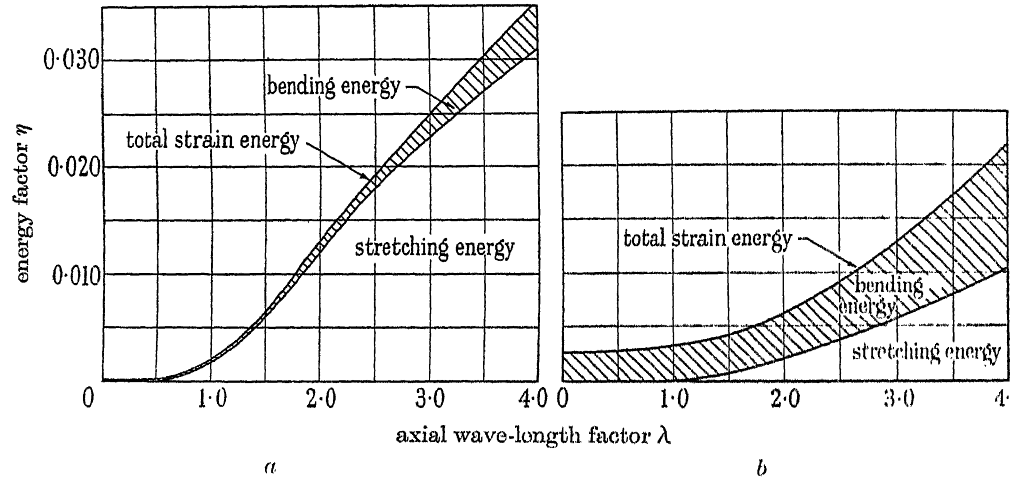


FIGURE 6. Strain energy due to bending and stretching ($\alpha = 0{\cdot}0525$. a, $n = 2$; b, $n = 4$).

Calculations of strain energy were also made for the thin experimental cylinder ($\alpha = 0{\cdot}01$), the results being plotted in figure 7. In this case, the stretching energy decreases rapidly with the increase of circumferential nodes, while the bending energy varies in the reverse manner. This results in a curve for total strain energy of somewhat parabolic form having a minimum value at $n = 7$. It will be observed that this curve corresponds in shape to the frequency curve given in figure 5 for freely supported ends. The curve for bending energy (figure 7) is also useful for comparison with the curve for free ends given in figure 5. In the latter case, Rayleigh's theory of inextension is applicable. With freely supported ends, however, the energy due to stretching is extremely small when n is greater than 10, and in this region the vibration approximates to one of inextension. The frequencies for both end-conditions are thus practically equal when n is large, as may be seen from figure 5.

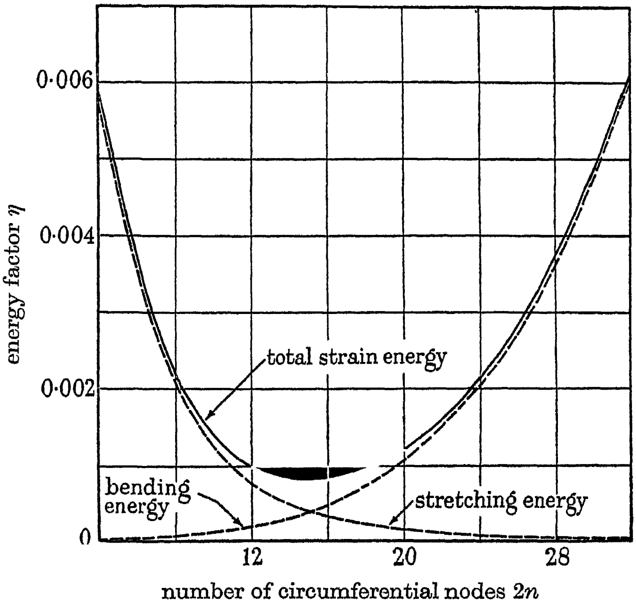


FIGURE 7. Strain energy due to bending and stretching ($\alpha = 0{\cdot}01$; $h = 0{\cdot}025$ in.; $a = 2{\cdot}51$ in.; $l = 2{\cdot}065$ in.).

As previously stated the cubic equation resulting from the mathematical analysis has, in general, three real positive roots, and thus for a given nodal configuration of the cylinder three frequencies are possible. In figure 8, calculated values of such frequencies are plotted for the experimental cylinder ($\alpha = 0{\cdot}0525$) for the case of eight circumferential nodes. The two higher frequencies, f_2 and f_3, lie considerably beyond the aural range, whatever value of n be considered. For $n = 4$, it will be seen that f_2 has a minimum value of 41,500 c./sec. Each curve of the diagram represents a particular type of vibration, but the nodal arrangements for each are identical at the same value of λ. The particular form of these vibrations may be appreciated by considering the relative magnitudes of the maximum component amplitudes in the axial, circumferential and radial directions. These may be calculated from equations (17) and (18) of appendix 3, and a series of values are given in table 2 for the case considered in figure 8.

TABLE 2. COMPONENT AMPLITUDE RATIOS ($\alpha = 0{\cdot}0525$; $n = 4$)

	axial wave-length factor λ							
	0·3867		1·547		2·707		3·867	
frequency	A/C	B/C	A/C	B/C	A/C	B/C	A/C	B/C
f_1	0·024	0·252	0·072	0·258	0·073	0·246	0·052	0·220
f_2	27·5	1·82	6·75	2·14	4·53	3·02	4·10	3·66
f_3	0·36	4·00	1·60	4·32	3·28	5·05	5·79	5·96

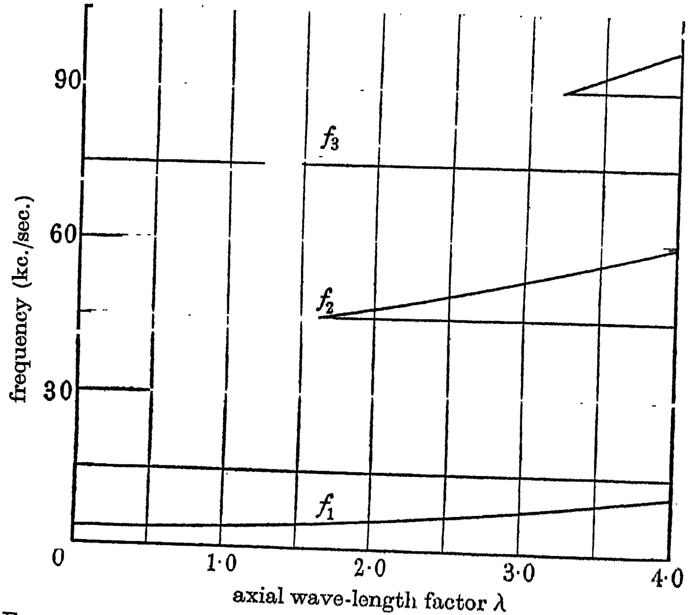


FIGURE 8. Calculated frequencies for similar nodal arrangements ($h = 0{\cdot}101$ in.; $a = 1{\cdot}925$ in.; $n = 4$).

It will be seen that for the lowest frequency f_1, which has been studied at length in this paper, the predominant amplitude is C; this means that the vibratory motion is mainly radial. The circumferential amplitude B is approximately $C/4$ for all values of λ, and the axial amplitude A has a maximum value of $C/14$. The former value corresponds to the case of free ends in which the ratio of circumferential to radial movement is given as $1/n$ (Rayleigh 1894, p. 386).

The vibration associated with f_2 is comprised mainly of axial motion when λ is small, but for larger values of λ the axial and circumferential components become almost equal. There is little bending of the cylinder walls in this case.

The highest frequency f_3 is shown to have a vibration which is predominantly circumferential, except at high values of λ when the axial motion becomes comparable.

The above results indicate that f_2 and f_3 are associated almost entirely with extensional vibration. In such cases, the strain energy is large, and this results in extremely high frequencies. It is this large energy contribution of stretching which is responsible for the strange variation of f_1 obtained from cylinders with freely supported ends. The higher frequencies have yet to be observed experimentally, but as this involves rather elaborate apparatus, such work must be left to the future.

APPENDIX 1

List of symbols

a	mean radius of cylinder.
A, B, C	maximum amplitudes of component vibrations.
e_x, e_y, e_z	direct strains.
e_{xy}, e_{yz}, e_{zx}	shear strains.
E	Young's modulus.
$f = \omega/2\pi$	frequency.
f_1, f_2, f_3	frequencies of similar nodal patterns.
g	acceleration due to gravity.
h	thickness of cylinder.
K_0, K_1, K_2	coefficients in frequency equation.
l	length of cylinder.
m	number of axial half-waves.
n	number of circumferential waves.
p_x, p_y, p_z	direct stresses.
p_{xy}, p_{yz}, p_{zx}	shear stresses.
S	total strain energy at any displacement.
S_b	strain energy due to bending.
S_s	strain energy due to stretching.
S_m	total strain energy at maximum displacement.
t	time.
T	total kinetic energy at any displacement.
u, v, w	component displacements of a point on the middle surface.
$U = A\cos\omega t$.	
$V = B\cos\omega t$.	
$W = C\cos\omega t$.	
x, y, z	co-ordinate distances in axial, circumferential and radial directions.
$\alpha = h/a$.	
$\beta = h^2/12a^2$.	
γ	shear strain in middle surface.
$\Delta = \dfrac{\rho a^2(1-\sigma^2)\,4\pi^2 f^2}{Eg}$	frequency factor.
ϵ_1, ϵ_2	direct strain in middle surface in directions of x and y.
η_b	energy factor (bending).
η_s	energy factor (stretching).
κ_1, κ_2	change of curvature of middle surface in directions of x and y.
$\lambda = \dfrac{\pi m a}{l}$	axial wave-length factor.
ρ	density.
σ	Poisson's ratio.
ϕ	angular co-ordinate.
τ	twist of middle surface.
$\omega = 2\pi f$	circular frequency.

APPENDIX 2

Derivation of frequency equation for thin cylindrical shells with freely supported ends

Consider a cylindrical shell of length l, thickness h and mean radius a, an element of which is shown in figure 9.

The element is bounded by two parallel planes which are perpendicular to the axis and distance δx apart, and by two radial planes which intersect at angle $\delta\phi$. The direct stresses acting on the element parallel to the X, Y and Z axes will be denoted respectively by p_x, p_y and p_z, and the shear stress acting on the face perpendicular to the X-axis in direction Y by p_{xy}. Shear stresses p_{yz}, p_{zx}, p_{yx}, ... have similar definitions. The direct strains corresponding to the above will be denoted by e_x, e_y and e_z and the shear strains by e_{xy}, e_{yz}, e_{zx}, ...

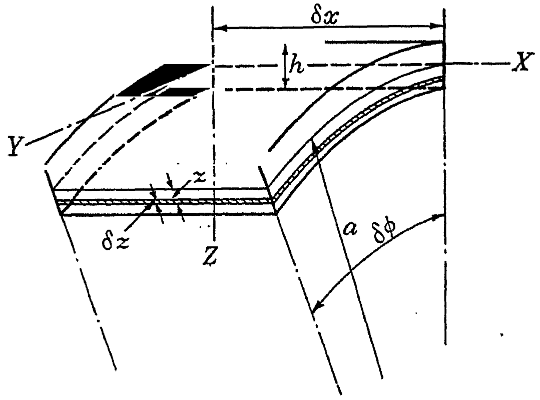


FIGURE 9. Element of thin cylinder.

To the first approximation the direct stress p_z and the shear strains e_{yz} and e_{zx} are zero (Love 1927, p. 529). The total strain energy for the deformed cylinder may thus be written, neglecting the trapezoidal form of the faces perpendicular to the X-axis, as

$$S = \int_0^{2\pi}\int_0^l\int_{-\frac{1}{2}h}^{+\frac{1}{2}h} \tfrac{1}{2}[p_x e_x + p_y e_y + p_{xy} e_{xy}]\, a\, d\phi\, dx\, dz. \tag{1}$$

Now from Hooke's law

$$p_x = \frac{E}{1-\sigma^2}(e_x+\sigma e_y), \quad p_y = \frac{E}{1-\sigma^2}(e_y+\sigma e_x), \quad p_{xy} = \frac{E}{2(1+\sigma)}e_{xy}, \tag{2}$$

which leads to

$$S = \frac{E}{2(1-\sigma^2)}\int_0^{2\pi}\int_0^l\int_{-\frac{1}{2}h}^{+\frac{1}{2}h}\left[e_x^2+e_y^2+2\sigma e_x e_y+\left(\frac{1-\sigma}{2}\right)e_{xy}^2\right] a\, d\phi\, dx\, dz. \tag{3}$$

Let the following symbols refer to the middle surface:

$e_1 e_2$	strains in directions x and y.
$\kappa_1 \kappa_2$	changes of curvature in directions x and y.
γ	shear strain.
τ	twist.

It can be shown that the strains at a distance z from the middle surface of a deformed shell can be expressed approximately (Love 1927, p. 529) as

$$e_x = \epsilon_1 - z\kappa_1, \quad e_y = \epsilon_2 - z\kappa_2, \quad e_{xy} = \gamma - 2z\tau. \tag{4}$$

If u, v and w are the displacements in the directions X, Y and Z of a point on the middle surface at any instant, then the strains and changes of curvature are given (Love 1927, p. 543) in terms of the displacements and their derivatives by

$$\left.\begin{aligned} \epsilon_1 &= \frac{\partial u}{\partial x}, & \epsilon_2 &= \frac{1}{a}\frac{\partial v}{\partial \phi} - \frac{w}{a}, & \gamma &= \frac{\partial v}{\partial x} + \frac{1}{a}\frac{\partial u}{\partial \phi}, \\ \kappa_1 &= \frac{\partial^2 w}{\partial x^2}, & \kappa_2 &= \frac{1}{a^2}\frac{\partial^2 w}{\partial \phi^2} + \frac{1}{a^2}\frac{\partial v}{\partial \phi}, & \tau &= \frac{1}{a}\frac{\partial^2 w}{\partial x \partial \phi} + \frac{1}{a}\frac{\partial v}{\partial x}, \end{aligned}\right\} \tag{5}$$

where ϕ defines the angular position of the point considered. Let

$$u = U\cos n\phi \cos\frac{m\pi x}{l}, \quad v = V\sin n\phi \sin\frac{m\pi x}{l}, \quad w = W\cos n\phi \sin\frac{m\pi x}{l}, \tag{6}$$

where n and $\frac{1}{2}m$ are respectively the number of circumferential and axial wave-lengths, and U, V and W are functions of time only. The above expressions satisfy the condition that the ends of the cylinder remain circular during vibration.

By expressing the strains in terms of U, V and W and substituting in (3), the integral may be evaluated to give

$$\begin{aligned} S = \frac{\pi E h l}{4a(1-\sigma^2)}\Big[& U^2\lambda^2 + (nV - W)^2 + 2\sigma\lambda U(W - nV) + \left(\frac{1-\sigma}{2}\right)(\lambda V - nU)^2 \\ & + \frac{h^2}{12a^2}\{\lambda^4 W^2 + (nV - n^2W)^2 + 2\sigma\lambda^2 W(n^2W - nV) + 2(1-\sigma)(\lambda V - \lambda n W)^2\}\Big], \end{aligned} \tag{7}$$

where $\lambda = \dfrac{m\pi a}{l}$.

The kinetic energy at any instant is given by

$$T = \frac{\rho}{2g}\int_0^{2\pi}\int_0^l\int_{-\frac{1}{2}h}^{+\frac{1}{2}h}\left[\left(\frac{\partial u}{\partial t}\right)^2 + \left(\frac{\partial v}{\partial t}\right)^2 + \left(\frac{\partial w}{\partial t}\right)^2\right] a\,d\phi\,dx\,dz$$

or

$$T = \frac{\pi\rho h l a}{4g}[\dot{U}^2 + \dot{V}^2 + \dot{W}^2]. \tag{8}$$

Since U, V and W are independent variables, the Lagrange equation is applicable. This gives

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{U}}\right) - \frac{\partial T}{\partial U} = -\frac{\partial S}{\partial U}, \tag{9}$$

and two similar equations in V and W.

From (7), (8) and (9)

$$\frac{\pi\rho h l a}{2g}\ddot{U} - 0 = -\frac{\pi E h l}{2a(1-\sigma^2)}\left[U\lambda^2 + \sigma\lambda(W - nV) - n\left(\frac{1-\sigma}{2}\right)(\lambda V - nU)\right]. \tag{10}$$

U, V and W are periodic with respect to time and may be written

$$U = A\cos\omega t, \quad V = B\cos\omega t, \quad W = C\cos\omega t, \tag{11}$$

where A, B and C are constants and $\omega/2\pi$ is the natural frequency of the vibration.

Hence
$$\left[\lambda^2+\left(\frac{1-\sigma}{2}\right)n^2-\Delta\right]A-\left(\frac{1+\sigma}{2}\right)\lambda nB+\sigma\lambda C = 0, \tag{12}$$

where
$$\Delta = \frac{\rho a^2(1-\sigma^2)\,\omega^2}{Eg}.$$

Similarly, by substituting in the Lagrange equations for V and W

$$-\left(\frac{1+\sigma}{2}\right)\lambda nA+\left[\left(\frac{1-\sigma}{2}\right)\lambda^2+n^2-\Delta+\beta\{n^2+2(1-\sigma)\,\lambda^2\}\right]B$$
$$-[n+\beta\{n^3+(2-\sigma)\,\lambda^2 n\}]\,C = 0 \tag{13}$$

and
$$\sigma\lambda A-[n+\beta\{n^3+(2-\sigma)\,\lambda^2 n\}]\,B+[1-\Delta+\beta(\lambda^2+n^2)^2]\,C = 0, \tag{14}$$

where $\beta = h^2/12a^2$.

Eliminating A, B and C from equations (12) to (14) leads to a cubic equation in Δ, the roots of which define the natural frequencies of the vibration. The equation is

$$\Delta^3-K_2\Delta^2+K_1\Delta-K_0 = 0, \tag{15}$$

where

$$K_0 = \tfrac{1}{2}(1-\sigma)^2(1+\sigma)\,\lambda^4+\tfrac{1}{2}(1-\sigma)$$
$$\times\beta[(\lambda^2+n^2)^4-2(4-\sigma^2)\,\lambda^4 n^2-8\lambda^2 n^4-2n^6+4(1-\sigma^2)\,\lambda^4+4\lambda^2 n^2+n^4],$$

$$K_1 = \tfrac{1}{2}(1-\sigma)\,(\lambda^2+n^2)^2+\tfrac{1}{2}(3-\sigma-2\sigma^2)\,\lambda^2+\tfrac{1}{2}(1-\sigma)\,n^2$$
$$+\beta[\tfrac{1}{2}(3-\sigma)\,(\lambda^2+n^2)^3+2(1-\sigma)\,\lambda^4-(2-\sigma^2)\,\lambda^2 n^2-\tfrac{1}{2}(3+\sigma)\,n^4+2(1-\sigma)\,\lambda^2+n^2],$$

and

$$K_2 = 1+\tfrac{1}{2}(3-\sigma)\,(\lambda^2+n^2)+\beta[(\lambda^2+n^2)^2+2(1-\sigma)\,\lambda^2+n^2].$$

APPENDIX 3

Strain energy due to bending and stretching of a cylindrical shell

By substituting the values for U, V and W of (11) in the expression for strain energy (7), the maximum strain energy may be written in terms of the component amplitudes

$$S_m = \frac{\pi Ehl}{4a(1-\sigma^2)}\left[A^2\lambda^2+(C-nB)^2+2\sigma\lambda A(C-nB)+\left(\frac{1-\sigma}{2}\right)(\lambda B-nA)^2\right.$$
$$\left.+\beta\{\lambda^4C^2+(n^2C-nB)^2+2\sigma\lambda^2C(n^2C-nB)+2(1-\sigma)\,(\lambda B-\lambda nC)^2\}\right]. \tag{16}$$

In this, the first four terms give the strain energy due to stretching and the remainder (dependent on ratio h/a) that due to bending. To determine what each effect contributes to the total energy, the ratio $A:B:C$ is required. By elimination between (12), (13) and (14)

$$\frac{A}{C}=\frac{\begin{array}{r}(1-\sigma)^2\lambda n^2-(1+\sigma)^2\lambda n^2\Delta-4\sigma^2\lambda\Delta+2\sigma^2(1-\sigma)\lambda^3+\beta[(1+\sigma)^2\lambda n^2(\lambda^2+n^2)^2\\ -4\sigma(1+\sigma)\lambda n^4-4\sigma(2-\sigma)(1+\sigma)\lambda^3n^2+4\sigma^2\lambda n^2+8\sigma^2(1-\sigma)\lambda^3]\end{array}}{\begin{array}{r}[2\lambda^2-2\Delta+(1-\sigma)n^2][(1-\sigma)n^2+2\sigma\Delta-\sigma(1-\sigma)\lambda^2\\ +\beta\{(1+\sigma)n^4+(2-\sigma)(1+\sigma)\lambda^2n^2-2\sigma n^2-4\sigma(1-\sigma)\lambda^2\}]\end{array}} \tag{17}$$

and

$$\frac{B}{C}=\frac{(1-\sigma)n-(1+\sigma)\Delta n+\beta[(1+\sigma)n(\lambda^2+n^2)^2-2\sigma(2-\sigma)\lambda^2n-2\sigma n^3]}{\begin{array}{r}(1-\sigma)n^2+2\sigma\Delta-\sigma(1-\sigma)\lambda^2+\beta[(1+\sigma)n^4\\ +(2-\sigma)(1+\sigma)\lambda^2n^2-2\sigma n^2-4\sigma(1-\sigma)\lambda^2]\end{array}}. \tag{18}$$

These results may be used to evaluate the stretching energy

$$S_s=\left[\frac{\pi EhlC^2}{4a(1-\sigma^2)}\right]\left[\lambda^2\left(\frac{A}{C}\right)^2+\left(1-n\frac{B}{C}\right)^2+2\sigma\lambda\left(\frac{A}{C}\right)\left(1-n\frac{B}{C}\right)+\left(\frac{1-\sigma}{2}\right)\left(\lambda\frac{B}{C}-n\frac{A}{C}\right)^2\right], \tag{19}$$

and the bending energy

$$S_b=\beta\left[\frac{\pi EhlC^2}{4a(1-\sigma^2)}\right]\left[(\lambda^2+n^2)^2+\{n^2+2(1-\sigma)\lambda^2\}\left(\frac{B}{C}\right)^2-2\{(2-\sigma)\lambda^2n+n^3\}\left(\frac{B}{C}\right)\right], \tag{20}$$

for any particular case in which Δ (proportional to square of frequency) has previously been obtained from (15). Equations (19) and (20) can be rewritten in the form

$$S_s=\eta_s\left[\frac{E\pi}{4(1-\sigma^2)}\right]lC^2 \quad\text{and}\quad S_b=\eta_b\left[\frac{E\pi}{4(1-\sigma^2)}\right]lC^2, \tag{21}$$

where η_s and η_b are non-dimensional energy factors for stretching and bending respectively, and the term in the bracket is a constant for the material.

APPENDIX 4

Note regarding the influence of approximations in the expressions for strain

The frequency equation derived by considering equilibrium is found to differ slightly from that obtained by the energy method depending on the accuracy of the expressions for strain. For example, if inertia terms are introduced into the displacement equations of Timoshenko (1940, p. 440) and expressions (6) are substituted for u, v and w, three equations are obtained corresponding to (12), (13) and (14). Comparison shows, however, that while two pairs are identical, that corresponding to (13) differs in the coefficients of two small terms. This divergence does not occur if the strain expressions given by Flügge (1934) are used, presumably because he does not approximate in deriving the strain relations. In general, any strain approximation results in divergences between the equations derived by the equilibrium and energy methods.

If the coefficients K_0, K_1 and K_2 of (15) be examined, it will be seen that each consists of two parts, one dependent on β, the other independent of β. In each case, the former contains one term of high power, e.g. $(\lambda^2+n^2)^4$ in the expression for K_0. In deriving K_0, K_1 and K_2 by the equilibrium method, using the strain expressions derived by Timoshenko (1940, p. 354), it is found that some of the terms of lower order dependent on β differ from those of (15), but the term of highest power and those independent of β are identical. In general, whatever strain expressions are used, the answers derived by the two methods differ only in the coefficients of some of the smaller terms. Moreover, these coefficients are only slightly different numerically, when a practical value for σ is introduced.

For the field covered by the present paper, namely, $h/a = 0$ to $0{\cdot}1$, $\lambda = 0$ to 4 and $n = 2$ to 7, the calculated frequencies are found to be practically independent of the strain approximations and method of derivation of the frequency equation. The greatest divergence in frequency recorded by using different strain expressions is only 0·7 %, and for most modes of vibration the difference is considerably less.

REFERENCES

Flügge, W. 1934 *Statik und Dynamik der Schalen.* Berlin: Springer.
Love, A. E. H. 1927 *Mathematical theory of elasticity,* 4th ed. Cambridge.
Rayleigh, Lord 1894 *Theory of sound,* 2nd ed. London: Macmillan.
Timoshenko, S. 1940 *Theory of plates and shells,* 1st ed. New York: McGraw-Hill.

On the quadratic form x^2-7y^2

By P. Varnavides
University College, London

(*Communicated by H. Davenport, F.R.S.—Received* 27 *October* 1948)

1. Let $f(x,y) = ax^2+bxy+cy^2$ be an indefinite binary quadratic form, with real coefficients a, b, c, and discriminant $d = b^2-4ac$. A well-known theorem of Minkowski asserts that for any real numbers x_0, y_0 there are integers x, y such that

$$|f(x+x_0, y+y_0)| \leqslant \tfrac{1}{4}\sqrt{d}. \tag{1}$$

Various authors (Heinhold 1938, 1939; Davenport 1946; Varnavides 1948*a*) have given results which are more precise than this. In particular, their results give, for a few special quadratic forms, the exact value of the constant, i.e. the least number by which $\frac{1}{4}\sqrt{d}$ may be replaced on the right of (1).

Perhaps the simplest form for which the existing results do not determine the exact value of the constant is the form x^2-7y^2, with discriminant 28. Here $\frac{1}{4}\sqrt{d} = \frac{1}{2}\sqrt{7}$, and this was improved by Heinhold (1938, p. 683) to $\frac{3}{4}$. In this paper, we attack the problem by a method based on that of Davenport (1947), and find the exact value of the constant. We prove that it is always possible to satisfy

$$|(x+x_0)^2-7(y+y_0)^2| \leqslant \tfrac{9}{14},$$

and that this would cease to be true if any lesser number stood on the right.

The problem can also be expressed in terms of the quadratic field $k(\sqrt{7})$ generated by $\sqrt{7}$. The general integer of this field is $\xi = x+y\sqrt{7}$, where x and y are rational integers, and its conjugate is $\xi' = x-y\sqrt{7}$. The result stated above can be expressed in the form: for any real numbers a, b there exists an integer ξ of $k(\sqrt{7})$ such that

$$|(\xi-a)(\xi'-b)| \leqslant \tfrac{9}{14}. \tag{2}$$

Our main theorem, which asserts rather more than this, is as follows:

THEOREM. *For any real numbers a, b there is an integer ξ of $k(\sqrt{7})$ which satisfies* (2). *If a, b are of the special form*

$$a = \frac{\omega}{\alpha_0}+\xi_0, \quad b = \frac{\omega'}{\alpha_0'}+\xi_0', \tag{3}$$

where

$$\alpha_0 = \tfrac{1}{9}(7+5\sqrt{7}), \quad \alpha_0' = \tfrac{1}{9}(7-5\sqrt{7}), \tag{4}$$

and ω is any unit of $k(\sqrt{7})$ and ξ_0 is any integer of $k(\sqrt{7})$, there are an infinity of solutions of (2) *with equality, but no solutions with strict inequality. If a, b are not of this special form, there is an integer ξ of $k(\sqrt{7})$ which satisfies the more precise inequality*

$$|(\xi-a)(\xi'-b)| \leqslant \frac{1}{1\cdot 56} < \frac{9}{14}. \tag{5}$$

The additional assertions in this theorem tell us that the result (2) is, in a sense, isolated. An interesting feature about the proof is that, although there is only one 'critical case', given by (3), this case arises in the course of the proof in two distinct ways.

In my thesis (1948*c*), I have obtained a similar result for the form x^2-11y^2, or the field $k(\sqrt{11})$. This is that for any real a, b there is an integer ξ of $k(\sqrt{11})$ such that

$$|(\xi-a)(\xi'-b)| \leqslant \tfrac{19}{22}.$$

The constant here is again the best possible, and the result is again isolated. The 'critical case' has the same form as above, but with

$$\alpha_0 = \tfrac{1}{19}(11+7\sqrt{11}).$$

2. *Proof of the theorem.* We first give a proof of the theorem based on a series of lemmas, which will be stated and proved later.

Let a, b be any real numbers. Let $M = M(a,b)$ denote the lower bound of $|(\xi-a)(\xi'-b)|$ for all integers ξ of the field $k(\sqrt{7})$. In proving the theorem, we may assume that

$$M \geqslant \frac{1}{1\cdot 56}, \tag{6}$$

and have then to prove that a and b must have the special form (3). The clauses of the theorem which relate to numbers of this special form are proved separately as lemma 10.

For any positive number ϵ_0, there exists an integer ξ_0 of the field such that

$$|(\xi_0-a)(\xi_0'-b)| = \frac{M}{1-\epsilon}, \tag{7}$$

where $0 \leqslant \epsilon < \epsilon_0$. We suppose throughout that ϵ_0 is sufficiently small. We define α, β by

$$\alpha = (a-\xi_0)^{-1}, \quad \beta = (b-\xi_0')^{-1}. \tag{8}$$

Then, using (6),

$$|\alpha\beta| = \frac{1-\epsilon}{M} \leqslant 1{\cdot}56. \tag{9}$$

Further, by the definition of M and by (7) and (8), we have

$$|(\alpha\xi-1)(\beta\xi'-1)| \geqslant 1-\epsilon \tag{10}$$

for all integers ξ of the field. If $\xi \neq 0$, this inequality can be written in the form

$$|(\alpha-\xi^{-1})(\beta-(\xi')^{-1})| \geqslant (1-\epsilon)\,|\xi\xi'|^{-1}. \tag{11}$$

We prove that if (9), (10) and (11) hold, then α, β must be of certain special forms. Lemma 1 shows that by means of certain transformations of α and β it is enough to consider values of α and β which satisfy

$$\alpha \geqslant \sqrt{(2-\epsilon)}, \tag{12}$$

$$\frac{\alpha}{\tau} \leqslant |\beta| \leqslant \alpha, \tag{13}$$

where τ is the fundamental unit of the field. In lemmas 2 to 9 we use (10) or (11), with various integers ξ of the field, to prove that if (9), (12) and (13) are satisfied, then *either* $\alpha = \alpha_0$ and $\beta = \alpha_0'$ *or* $\alpha = \alpha_1$ and $\beta = \alpha_1'$, where

$$\alpha_1 = \tfrac{1}{3}(7+3\sqrt{7}).$$

It then follows from lemma 1 that in the general case, when α and β are not restricted by (12) and (13), they must have one of the following four forms:

$$\alpha = \omega\alpha_0, \quad \beta = \omega'\alpha_0'; \qquad \alpha = \omega\alpha_0', \quad \beta = \omega'\alpha_0;$$
$$\alpha = \omega\alpha_1, \quad \beta = \omega'\alpha_1'; \qquad \alpha = \omega\alpha_1', \quad \beta = \omega'\alpha_1.$$

It now follows from (8) that a, b have one of the four forms

$$\left.\begin{aligned} a &= \frac{\omega}{\alpha_0}+\xi_0, \quad b = \frac{\omega'}{\alpha_0'}+\xi_0',\\ a &= \frac{\omega}{\alpha_0'}+\xi_0, \quad b = \frac{\omega'}{\alpha_0}+\xi_0',\\ a &= \frac{\omega}{\alpha_1}+\xi_0, \quad b = \frac{\omega'}{\alpha_1'}+\xi_0',\\ a &= \frac{\omega}{\alpha_1'}+\xi_0, \quad b = \frac{\omega'}{\alpha_1}+\xi_0', \end{aligned}\right\} \tag{14}$$

where ω is some unit and ξ_0 some integer of the field. But it is easily verified that

$$\frac{1}{\alpha_0'}=-\frac{1}{\alpha_0}-1,\quad \frac{1}{\alpha_1}=-\frac{1}{\alpha_0}-2+\sqrt{7},\quad \frac{1}{\alpha_1'}=\frac{1}{\alpha_0}-1-\sqrt{7}. \tag{15}$$

Consequently, if a, b have any one of the four forms in (14), they are also expressible in the form (3). This completes the proof of the theorem.

3. LEMMA 1. *Let $k(\sqrt{m})$ be any real quadratic field. Let α, β be real numbers such that $\alpha\beta \neq 0$ and such that*

$$|(\alpha\xi-1)(\beta\xi'-1)| \geqslant 1-\epsilon$$

for every integer ξ of $k(\sqrt{m})$. Then there exists a unit ω of $k(\sqrt{m})$ such that the numbers α^*, β^* defined by one of the two substitutions

$$\alpha=\omega\alpha^*,\quad \beta=\omega'\beta^*;\qquad \alpha=\omega'\beta^*,\quad \beta=\omega\alpha^*,$$

also have the above property, and in addition satisfy

$$|\alpha^*\beta^*| = |\alpha\beta|,\quad \alpha^* \geqslant \sqrt{(2-\epsilon)},\quad \frac{\alpha^*}{\tau} \leqslant |\beta^*| \leqslant \alpha^*,$$

where τ is the fundamental unit of the field.

This is lemma 1 of Varnavides (1948*b*).

4. In the following lemmas 2 to 9, we suppose that (9), (12) and (13) are satisfied, and that (10) and (11) are satisfied for every integer ξ in the field $k(\sqrt{7})$, other than 0. We shall prove that either $\alpha=\alpha_0$, $\beta=\beta_0$ or $\alpha=\alpha_1$, $\beta=\beta_1$.

It is convenient to tabulate for reference the numerical values of certain integers of $k(\sqrt{7})$ and their conjugates, and the values of α_0, α_0', α_1, α_1' (see table 1). We note also that the fundamental unit τ of $k(\sqrt{7})$ is given by

$$\tau = 8+3\sqrt{7} = \tfrac{1}{2}(3+\sqrt{7})^2 = 15\cdot 93725\ldots. \tag{16}$$

TABLE 1. SOME INTEGERS OF $k(\sqrt{7})$

n	ξ_n	$1/\xi_n$	$1/\xi_n'$	$\xi_n\xi_n'$
1	-1	-1	-1	1
2	$1+\sqrt{7}$	$\tfrac{1}{6}(-1+\sqrt{7}) = 0\cdot 27429\ldots$	$-\tfrac{1}{6}(1+\sqrt{7}) = -0\cdot 60762\ldots$	-6
3	-2	$-\tfrac{1}{2}$	$-\tfrac{1}{2}$	4
4	-3	$-\tfrac{1}{3}$	$-\tfrac{1}{3}$	9
5	$-1+\sqrt{7}$	$\tfrac{1}{6}(1+\sqrt{7}) = 0\cdot 60762\ldots$	$-\tfrac{1}{6}(-1+\sqrt{7}) = -0\cdot 27429\ldots$	-6
6	$-2+\sqrt{7}$	$\tfrac{1}{3}(2+\sqrt{7}) = 1\cdot 54858\ldots$	$-\tfrac{1}{3}(-2+\sqrt{7}) = -0\cdot 21525\ldots$	-3
7	$3-\sqrt{7}$	$\tfrac{1}{2}(3+\sqrt{7}) = 2\cdot 82287\ldots$	$\tfrac{1}{2}(3-\sqrt{7}) = 0\cdot 17712\ldots$	2
8	$-5+2\sqrt{7}$	$\tfrac{1}{3}(5+2\sqrt{7}) = 3\cdot 43050\ldots$	$-\tfrac{1}{3}(-5+2\sqrt{7}) = -0\cdot 09716\ldots$	-3
9	$-13+5\sqrt{7}$	$\tfrac{1}{6}(13+5\sqrt{7}) = 4\cdot 37145\ldots$	$-\tfrac{1}{6}(-13+5\sqrt{7}) = -0\cdot 03812\ldots$	-6
10	$24-9\sqrt{7}$	$\tfrac{1}{3}(8+3\sqrt{7}) = 5\cdot 31241\ldots$	$\tfrac{1}{3}(8-3\sqrt{7}) = 0\cdot 02091\ldots$	9

$$\alpha_0 = \tfrac{1}{9}(7+5\sqrt{7}) = 2\cdot 24763\ldots \qquad \alpha_0' = -\tfrac{1}{9}(-7+5\sqrt{7}) = -0\cdot 69208\ldots$$
$$\alpha_1 = \tfrac{1}{3}(7+3\sqrt{7}) = 4\cdot 97908\ldots \qquad \alpha_1' = -\tfrac{1}{3}(-7+3\sqrt{7}) = -0\cdot 31241\ldots$$

LEMMA 2. *We have*

$$1\cdot 33 < |\alpha\beta| \leqslant 1\cdot 56, \tag{17}$$

$$1\cdot 41 < \alpha < 4\cdot 99, \tag{18}$$

$$0\cdot 266 < -\beta < 1\cdot 11. \tag{19}$$

Proof. By (9),
$$|\alpha\beta| \leqslant 1{\cdot}56. \tag{20}$$
Also, by the definition of M and by Heinhold's result, quoted in § 1, we have $M \leqslant \frac{3}{4}$. Thus
$$|\alpha\beta| = \frac{1-\epsilon}{M} \geqslant \tfrac{4}{3}(1-\epsilon) > 1{\cdot}33. \tag{21}$$
This proves (17).

From (20) and (13),
$$\frac{\alpha^4}{\tau} \leqslant 1{\cdot}56,$$
so that by (16)
$$\alpha \leqslant \sqrt{\{1{\cdot}56\tau\}} < \sqrt{\{1{\cdot}56 \times 15{\cdot}94\}} < 4{\cdot}99. \tag{22}$$
Also from (12)
$$1{\cdot}41 < \sqrt{(2-\epsilon)} \leqslant \alpha. \tag{23}$$
These results prove (18).

By use of (18) in (17) we obtain
$$0{\cdot}266 < \frac{1{\cdot}33}{4{\cdot}99} < |\beta| < \frac{1{\cdot}56}{1{\cdot}41} < 1{\cdot}11. \tag{24}$$
Now, to prove (19), we have to prove that $\beta < 0$. We consider two cases.

(i) Suppose that $\beta > 0$ and $2{\cdot}8 \leqslant \alpha < 4{\cdot}99$. Then, by (18),
$$\begin{aligned} |1-\xi_7^{-1}\alpha^{-1}| &= |1-2{\cdot}82287\ldots\alpha^{-1}| \\ &< 1-\tfrac{1}{5}2{\cdot}82 < 0{\cdot}44. \end{aligned} \tag{25}$$
Also, using (17) and the hypothesis that $\beta > 0$ and $\alpha \geqslant 2{\cdot}8$,
$$\begin{aligned} |\alpha\beta-(\xi_7')^{-1}\alpha| &= |\alpha\beta - 0{\cdot}17712\ldots\alpha| \\ &< 1{\cdot}56 - 0{\cdot}177 \times 2{\cdot}8 \\ &< 1{\cdot}07. \end{aligned} \tag{26}$$
Multiplying (25) and (26), we obtain
$$|(\alpha-\xi_7^{-1})(\beta-(\xi_7')^{-1})| < 0{\cdot}48,$$
which is contrary to (11) with $\xi = \xi_7$.

(ii) Suppose that $\beta > 0$ and $1{\cdot}41 < \alpha < 2{\cdot}8$. Then, by (24) and (17),
$$1{\cdot}11 > \beta \geqslant \frac{1{\cdot}33}{\alpha} > \frac{1{\cdot}33}{2{\cdot}8} > 0{\cdot}47.$$
Thus
$$|\alpha-1| < 1{\cdot}8 \quad \text{and} \quad |\beta-1| < 0{\cdot}53.$$
Multiplying, we obtain
$$|(\alpha-1)(\beta-1)| < 0{\cdot}954,$$
contrary to (10) with $\xi = 1$.

These contradictions prove that $\beta < 0$ and complete the proof of the lemma.

LEMMA 3. *If* $\alpha \leqslant 3{\cdot}2$, *then*
$$1{\cdot}6 < \alpha < 2{\cdot}4, \tag{27}$$
$$0{\cdot}65 < -\beta < 0{\cdot}9. \tag{28}$$

Proof. We write $\bar{\beta} = -\beta$ so that $\bar{\beta} > 0$. Suppose that $\bar{\beta} \geqslant 0{\cdot}9$. Then, using (18) and (19) and our assumption that $\alpha \leqslant 3{\cdot}2$,
$$|\alpha+1| \leqslant 4{\cdot}2 \quad \text{and} \quad |\beta+1| = |\bar{\beta}-1| < 0{\cdot}11.$$

Multiplying, we obtain $\quad |(\alpha+1)(\beta+1)| < 0\cdot47,$

contrary to (11) with $\xi = -1$. Thus $\quad \bar\beta < 0\cdot9. \qquad (29)$

Now suppose that $\alpha \geqslant 2\cdot4$. Then, by (17),

$$\bar\beta \leqslant \frac{1\cdot56}{2\cdot4} = 0\cdot65.$$

Since $2\cdot4 \leqslant \alpha \leqslant 3\cdot2, \quad |\alpha-\xi_7^{-1}| = |\alpha-2\cdot82287\ldots| < 0\cdot5;$

and $\quad |\beta-(\xi_7')^{-1}| = |\bar\beta+0\cdot17712\ldots| < 0\cdot83.$

Multiplying, $\quad |(\alpha-\xi_7^{-1})(\beta-(\xi_7')^{-1})| < 0\cdot42,$

contrary to (11) with $\xi = \xi_7$. Hence $\quad \alpha < 2\cdot4. \qquad (30)$

Now suppose that $\alpha \leqslant 1\cdot6$. Then, by (18), (19) and (29),

$$|\alpha-\xi_6^{-1}| = |\alpha-1\cdot54858\ldots| < 0\cdot15,$$

$$|\beta-(\xi_6')^{-1}| = |\bar\beta-0\cdot21525\ldots| < 0\cdot7.$$

Multiplying, we obtain $\quad |(\alpha-\xi_6^{-1})(\beta-(\xi_6')^{-1})| < 0\cdot11,$

contrary to (11) with $\xi = \xi_6$, since $|\xi_6\xi_6'|^{-1} = \frac{1}{3}$. Thus (27) is proved.

By (17) and (27), $\quad \bar\beta \geqslant \dfrac{1\cdot33}{2\cdot4} > 0\cdot55.$

Suppose that $\bar\beta \leqslant 0\cdot65$. Then, using (27),

$$|\alpha-\xi_2^{-1}| = |\alpha-0\cdot27429\ldots| < 2\cdot2,$$

$$|\beta-(\xi_2')^{-1}| = |\bar\beta-0\cdot60762\ldots| < 0\cdot06,$$

so that $\quad |(\alpha-\xi_2^{-1})(\beta-(\xi_2')^{-1})| < 0\cdot14,$

contrary to (11) with $\xi = \xi_2$, since $|\xi_2\xi_2'|^{-1} = 0\cdot16666\ldots$. This completes the proof of the lemma.

Lemma 4. *We have*

$$N(\alpha_0\xi_r-1) = 1 \quad \textit{for } r = 1, 2, 6 \textit{ and } 7, \tag{31}$$

and

$$N(\alpha_1\xi_r-1) = 1 \quad \textit{for } r = 4, 5, 9 \textit{ and } 10. \tag{32}$$

Proof. Since

$$\alpha_0 = \tfrac{1}{9}(7+5\sqrt{7}) = \frac{2\sqrt{7}}{5-\sqrt{7}}, \tag{33}$$

$$\alpha_1 = \tfrac{1}{3}(7+3\sqrt{7}) = \frac{2\sqrt{7}}{9-3\sqrt{7}}, \tag{34}$$

we have to prove that

$$N(2\sqrt{7}\,\xi_r-5+\sqrt{7}) = N(5-\sqrt{7}) = 18$$

for $r = 1, 2, 6$ and 7 and that

$$N(2\sqrt{7}\,\xi_r-9+3\sqrt{7}) = N(9-3\sqrt{7}) = 18$$

for $r = 4, 5, 9$ and 10.

Now
$$\begin{aligned} N(2\sqrt{7}\,\xi_1 -5+\sqrt{7}) &= N(-5-\sqrt{7}) &&= 18,\\ N(2\sqrt{7}\,\xi_2 -5+\sqrt{7}) &= N(9+3\sqrt{7}) &&= 18,\\ N(2\sqrt{7}\,\xi_6 -5+\sqrt{7}) &= N(9-3\sqrt{7}) &&= 18,\\ N(2\sqrt{7}\,\xi_7 -5+\sqrt{7}) &= N(-19+7\sqrt{7}) &&= 18,\end{aligned}$$

and
$$\begin{aligned} N(2\sqrt{7}\,\xi_4 -9+3\sqrt{7}) &= N(-9-3\sqrt{7}) &&= 18,\\ N(2\sqrt{7}\,\xi_5 -9+3\sqrt{7}) &= N(5+\sqrt{7}) &&= 18,\\ N(2\sqrt{7}\,\xi_9 -9+3\sqrt{7}) &= N(61-23\sqrt{7}) &&= 18,\\ N(2\sqrt{7}\,\xi_{10} -9+3\sqrt{7}) &= N(-135+51\sqrt{7}) &&= 18.\end{aligned}$$

This proves the lemma.

LEMMA 5. *If $\alpha \leqslant 3{\cdot}2$, then for some absolute constant C,*

$$|\alpha-\alpha_0| \leqslant C\epsilon, \quad |\beta-\alpha_0'| \leqslant C\epsilon. \tag{35}$$

Proof. We write
$$\alpha = \alpha_0+\gamma, \quad \beta = \alpha_0'+\delta. \tag{36}$$

Then, by (10), for every integer ξ of $k(\sqrt{7})$, we have

$$|\{\alpha_0\xi-1+\gamma\xi\}\{\alpha_0'\xi'-1+\delta\xi'\}| \geqslant 1-\epsilon,$$

so that
$$\left|\left\{1+\frac{\xi}{\alpha_0\xi-1}\gamma\right\}\left\{1+\frac{\xi'}{\alpha_0'\xi'-1}\delta\right\}\right| \geqslant \frac{1-\epsilon}{|N(\alpha_0\xi-1)|}. \tag{37}$$

We take $\xi = \xi_r$ for $r = 1, 2, 6$ and 7. Then, by lemma 4,

$$|(1+\rho_r\gamma)(1+\sigma_r\delta)| \geqslant 1-\epsilon \tag{38}$$

for $r = 1, 2, 6$ and 7, where, using (33),

$$\rho_r = \frac{\xi_r}{\alpha_0\xi_r-1} = \frac{(5-\sqrt{7})\,\xi_r}{2\sqrt{7}\,\xi_r-5+\sqrt{7}},$$

$$\sigma_r = \frac{\xi_r'}{\alpha_0'\xi_r'-1} = \frac{(5+\sqrt{7})\,\xi_r'}{-2\sqrt{7}\,\xi_r'-5-\sqrt{7}},$$

for $r = 1, 2, 6$ and 7. Thus

$$\rho_1 = \frac{(5-\sqrt{7})(-1)}{-5-\sqrt{7}} > 0, \qquad \sigma_1 = \frac{(5+\sqrt{7})(-1)}{-5+\sqrt{7}} > 0,$$

$$\rho_2 = \frac{(5-\sqrt{7})(1+\sqrt{7})}{9+3\sqrt{7}} > 0, \quad \sigma_2 = \frac{(5+\sqrt{7})(1-\sqrt{7})}{9-3\sqrt{7}} < 0,$$

$$\rho_6 = \frac{(5-\sqrt{7})(\sqrt{7}-2)}{9-3\sqrt{7}} > 0, \quad \sigma_6 = \frac{(5+\sqrt{7})(-\sqrt{7}-2)}{9+3\sqrt{7}} < 0,$$

$$\rho_7 = \frac{(5-\sqrt{7})(3-\sqrt{7})}{-19+7\sqrt{7}} < 0, \quad \sigma_7 = \frac{(5+\sqrt{7})(3+\sqrt{7})}{-19-7\sqrt{7}} < 0.$$

Now
$$1+\rho_r\gamma = \frac{\alpha-(\xi_r)^{-1}}{\alpha_0-(\xi_r)^{-1}}, \quad 1+\sigma_r\delta = \frac{\beta-(\xi_r')^{-1}}{\alpha_0'-(\xi_r')^{-1}},$$

and it is clear from the inequalities (27) and (28) and the table that $\alpha-(\xi_r)^{-1}$ and $\alpha_0-(\xi_r)^{-1}$ and also $\beta-(\xi_r')^{-1}$ and $\alpha_0'-(\xi_r')^{-1}$ have the same signs, for $r = 1, 2, 6$ and 7. So $1+\rho_r\gamma$ and $1+\sigma_r\delta$ are positive and, by (38) and the inequality of the arithmetic and geometric means,

$$\tfrac{1}{2}\{1+\rho_r\gamma+1+\sigma_r\delta\}\geqslant 1-\epsilon,$$

i.e.

$$\rho_r\gamma+\sigma_r\delta\geqslant -2\epsilon, \quad \text{for} \quad r = 1, 2, 6 \text{ and } 7. \tag{39}$$

Eliminating first δ and then γ from the first and last of the inequalities (39), we obtain

$$\left.\begin{aligned}(\rho_1|\sigma_7|+\rho_7\sigma_1)\gamma&\geqslant -2\epsilon\{|\sigma_7|+\sigma_1\},\\ (\sigma_1|\rho_7|+\sigma_7\rho_1)\delta&\geqslant -2\epsilon\{|\rho_7|+\rho_1\}.\end{aligned}\right\} \tag{40}$$

Since

$$\begin{aligned}\rho_1|\sigma_7|+\rho_7\sigma_1 &= -\{\sigma_1|\rho_7|+\sigma_7\rho_1\}\\ &= \rho_1|\sigma_7|-|\rho_7|\sigma_1\\ &= \frac{(5-\sqrt{7})(3+\sqrt{7})}{19+7\sqrt{7}}-\frac{(5+\sqrt{7})(3-\sqrt{7})}{19-7\sqrt{7}}\\ &= \frac{(4+\sqrt{7})(3-\sqrt{7})(3+\sqrt{7})}{(4+\sqrt{7})(3+\sqrt{7})}-\frac{(4-\sqrt{7})(3+\sqrt{7})(3-\sqrt{7})}{(4-\sqrt{7})(3-\sqrt{7})}\\ &= -2\sqrt{7}<0,\end{aligned}$$

the inequalities (40) show that

$$\gamma\leqslant C_1\epsilon, \quad \delta\geqslant -C_2\epsilon,$$

for some absolute constants C_1 and C_2. Substituting these bounds for γ and δ in the second or third of the inequalities (39) we obtain $\delta\leqslant C_3\epsilon$ and $\gamma\geqslant -C_4\epsilon$ for some constants C_3 and C_4. It now follows from (36) that (35) is satisfied with

$$C = \max\{C_1, C_2, C_3, C_4\}.$$

LEMMA 6. *If* $\alpha\leqslant 3{\cdot}2$, *then* $\quad \alpha = \alpha_0, \quad \beta = \alpha_0'.$ (41)

Proof. Write

$$\eta_n = \frac{1-\tau^{2n}}{\alpha_0}, \quad \zeta_n = (-2+\sqrt{7})\tau^{2n}+\frac{1-\tau^{2n}}{\alpha_0}. \tag{42}$$

Then, as

$$\frac{1-\tau^{2n}}{\alpha_0} = (5-\sqrt{7})\tau^n\left\{\frac{\tau^{-n}-\tau^n}{2\sqrt{7}}\right\},$$

and

$$\tau^n\equiv\tau^{-n}\bmod(2\sqrt{7}),$$

for any rational integer n, it follows that η_n and ζ_n are integers of $k(\sqrt{7})$, when n is any rational integer. Further,

$$N(\alpha_0\eta_n-1) = N(-\tau^{2n}) = 1,$$

and

$$\begin{aligned}N(\alpha_0\zeta_n-1) &= N(\alpha_0\xi_6\tau^{2n}-\tau^{2n})\\ &= N(\alpha_0\xi_6-1) = 1.\end{aligned}$$

Thus, by (37),

$$\left|\left\{1+\frac{\eta_n}{\alpha_0\eta_n-1}\gamma\right\}\left\{1+\frac{\eta_n'}{\alpha_0'\eta_n'-1}\delta\right\}\right|\geqslant 1-\epsilon,$$

$$\left|\left\{1+\frac{\zeta_n}{\alpha_0\zeta_n-1}\gamma\right\}\left\{1+\frac{\zeta_n'}{\alpha_0'\zeta_n'-1}\delta\right\}\right|\geqslant 1-\epsilon,$$

for all integers n. Writing

$$\lambda_n = \frac{\eta_n}{\alpha_0 \eta_n - 1} = \frac{1-\tau^{-2n}}{\alpha_0}, \tag{43}$$

$$\mu_n = \frac{\zeta_n}{\alpha_0 \zeta_n - 1} = \frac{\alpha_0 \xi_6 \tau^{2n} - \tau^{2n} + 1}{\alpha_0\{\alpha_0 \xi_6 \tau^{2n} - \tau^{2n}\}}$$

$$= \frac{1+\tau^{-2n}\{\alpha_0 \xi_6 - 1\}^{-1}}{\alpha_0}, \tag{44}$$

we have

$$|(1+\lambda_n\gamma)(1+\lambda_n'\delta)| \geqslant 1-\epsilon, \tag{45}$$

$$|(1+\mu_n\gamma)(1+\mu_n'\delta)| \geqslant 1-\epsilon. \tag{46}$$

Now suppose that $\gamma > 0$. Then, if n is a negative rational integer,

$$|\lambda_n'| = \left|\frac{1-\tau^{2n}}{\alpha_0'}\right| < 1/|\alpha_0'| < 2.$$

Hence, by (35),

$$|1+\lambda_n'\delta| \leqslant 1+2C\epsilon, \tag{47}$$

if n is a negative rational integer. Since

$$|\lambda_{-1} - \gamma| \leqslant |\lambda_{-1}| C\epsilon < 1,$$

assuming that ϵ is sufficiently small, and

$$\frac{\lambda_{n-1}}{\lambda_n} = \frac{1-\tau^{-2n+2}}{1-\tau^{-2n}} = \frac{\tau^2-\tau^{2n}}{1-\tau^{2n}}$$

$$\leqslant \frac{\tau^2-\tau^{-2}}{1-\tau^{-2}} = \tau^2+1,$$

for $n \leqslant -1$, while

$$\lambda_n \to -\infty \quad \text{as} \quad n \to -\infty;$$

it is clear that we can choose a negative rational integer n such that

$$(1+\tau^2)^{-1} \leqslant -\lambda_n\gamma < 1.$$

For this value of n,

$$|1+\lambda_n\gamma| \leqslant 1-(1+\tau^2)^{-1}. \tag{48}$$

Now (47) and (48) are contrary to (45), provided our original choice of ϵ_0 was sufficiently small. Hence we must have $\gamma \leqslant 0$.

Similarly, we suppose that $\delta < 0$, and we consider positive values of n. Then λ_n is bounded and λ_n' tends to $-\infty$, the ratio $\lambda_{n+1}'/\lambda_n'$ being bounded, as $n \to +\infty$ Arguing as before we obtain a contradiction.

Further, as $n \to +\infty$, the sequence μ_n is bounded and $\mu_n' \to -\infty$, the ratio μ_{n+1}'/μ_n' being bounded; while as $n \to -\infty$ the sequence μ_n' is bounded and $\mu_n \to +\infty$, the ratio μ_{n-1}/μ_n being bounded. Using (46) in place of (45), we deduce that $\delta \leqslant 0$ and $\gamma \geqslant 0$. Hence $\gamma = \delta = 0$ and so, by (36), we have established (41).

Lemma 7. *If* $\alpha > 3{\cdot}2$, *then*

$$4{\cdot}75 < \alpha < 4{\cdot}99, \tag{49}$$

$$0{\cdot}29 < -\beta < 0{\cdot}33. \tag{50}$$

Proof. Suppose that $\alpha \leqslant 4$. Then $3{\cdot}2 < \alpha \leqslant 4$ and

$$|\alpha - \xi_8^{-1}| = |\alpha - 3{\cdot}43050\ldots| < 0{\cdot}6.$$

Also, by (19) and (17), $$0{\cdot}266 < \bar{\beta} \leqslant \frac{1{\cdot}56}{\alpha} < \frac{1{\cdot}56}{3{\cdot}2} < 0{\cdot}49,$$

so that $$|\beta - (\xi_8')^{-1}| = |\bar{\beta} - 0{\cdot}09716\ldots| < 0{\cdot}4.$$

Multiplying these results,

$$|(\alpha - \xi_8^{-1})(\beta - (\xi_8')^{-1})| < 0{\cdot}24,$$

contrary to (11) with $\xi = \xi_8$, since $|\xi_8 \xi_8'|^{-1} = 0{\cdot}333\ldots$. Hence $\alpha > 4$.

Now suppose that $\alpha \leqslant 4{\cdot}75$. Then, by (19) and (17),

$$0{\cdot}266 < \bar{\beta} \leqslant \frac{1{\cdot}56}{\alpha} < \frac{1{\cdot}56}{4} = 0{\cdot}39.$$

Thus $$|\alpha - \xi_9^{-1}| = |\alpha - 4{\cdot}37145\ldots| < 0{\cdot}4,$$

and $$|\beta - (\xi_9')^{-1}| = |\bar{\beta} - 0{\cdot}03812\ldots| < 0{\cdot}4.$$

Multiplying these results,

$$|(\alpha - \xi_9^{-1})(\beta - (\xi_9')^{-1})| < 0{\cdot}16,$$

contrary to (11) with $\xi = \xi_9$, since $|\xi_9 \xi_9'|^{-1} = 0{\cdot}1666\ldots$. Hence $\alpha > 4{\cdot}75$ and (49) follows from (18).

By (49) and (17), $$\bar{\beta} \leqslant \frac{1{\cdot}56}{\alpha} < \frac{1{\cdot}56}{4{\cdot}75} < 0{\cdot}33,$$

and by (49) and (12), $$\bar{\beta} \geqslant \frac{\alpha}{\tau} = \frac{\alpha}{15{\cdot}93725\ldots} > \frac{4{\cdot}75}{16} > 0{\cdot}29.$$

Thus $\beta = -\bar{\beta}$ satisfies (50).

LEMMA 8. *If $\alpha > 3{\cdot}2$, then for some absolute constant C',*

$$|\alpha - \alpha_1| \leqslant C'\epsilon, \quad |\beta - \alpha_1'| \leqslant C'\epsilon. \tag{51}$$

Proof. We write $$\alpha = \alpha_1 + \kappa, \quad \beta = \alpha_1' + \nu. \tag{52}$$

Then, by (10), for every integer ξ of $k(\sqrt{7})$, we have

$$|\{\alpha_1 \xi - 1 + \kappa\xi\}\{\alpha_1'\xi' - 1 + \nu\xi'\}| \geqslant 1 - \epsilon,$$

so that $$\left|\left\{1 + \frac{\xi}{\alpha_1\xi - 1}\kappa\right\}\left\{1 + \frac{\xi'}{\alpha_1'\xi' - 1}\nu\right\}\right| \geqslant \frac{1-\epsilon}{|N(\alpha_1\xi - 1)|}.$$

We take $\xi = \xi_r$, for $r = 4$, 5, 9 and 10. Then, by lemma 4,

$$|(1 + \rho_r\kappa)(1 + \sigma_r\nu)| \geqslant 1 - \epsilon \quad (r = 4, 5, 9 \text{ and } 10), \tag{53}$$

where, using (34),

$$\rho_r = \frac{\xi_r}{\alpha_1\xi_r - 1} = \frac{(9 - 3\sqrt{7})\xi_r}{2\sqrt{7}\,\xi_r - 9 + 3\sqrt{7}}, \quad \sigma_r = \frac{\xi_r'}{\alpha_1'\xi_r' - 1} = \frac{(9 + 3\sqrt{7})\xi_r'}{-2\sqrt{7}\,\xi_r' - 9 - 3\sqrt{7}},$$

for $r = 4, 5, 9$ and 10. Thus

$$\rho_4 = \frac{(9-3\sqrt{7})(-3)}{-9-3\sqrt{7}} > 0, \qquad \sigma_4 = \frac{(9+3\sqrt{7})(-3)}{-9+3\sqrt{7}} > 0,$$

$$\rho_5 = \frac{(9-3\sqrt{7})(-1+\sqrt{7})}{5+\sqrt{7}} > 0, \qquad \sigma_5 = \frac{(9+3\sqrt{7})(-1-\sqrt{7})}{5-\sqrt{7}} < 0,$$

$$\rho_9 = \frac{(9-3\sqrt{7})(-13+5\sqrt{7})}{61-23\sqrt{7}} > 0, \qquad \sigma_9 = \frac{(9+3\sqrt{7})(-13-5\sqrt{7})}{61+23\sqrt{7}} < 0,$$

$$\rho_{10} = \frac{(9-3\sqrt{7})(24-9\sqrt{7})}{-135+51\sqrt{7}} < 0, \qquad \sigma_{10} = \frac{(9+3\sqrt{7})(24+9\sqrt{7})}{-135-51\sqrt{7}} < 0.$$

Now
$$1+\rho_r\kappa = \frac{\alpha-(\xi_r)^{-1}}{\alpha_1-(\xi_r)^{-1}}, \quad 1+\sigma_r\nu = \frac{\beta-(\xi_r')^{-1}}{\alpha_1'-(\xi_r')^{-1}},$$

and it is clear from the inequalities (49) and (50) and table 1 that the numbers of the pairs $\alpha-(\xi_r)^{-1}$, $\alpha_1-(\xi_r)^{-1}$ and $\beta-(\xi_r')^{-1}$, $\alpha_1'-(\xi_r')^{-1}$ have the same signs, for $r = 4, 5, 9$ and 10. So $1+\rho_r\kappa$ and $1+\sigma_r\nu$ are positive and, by (53) and the inequality of the arithmetic and geometric means,

$$\tfrac{1}{2}\{1+\rho_r\kappa+1+\sigma_r\nu\} \geqslant 1-\epsilon,$$

i.e.
$$\rho_r\kappa+\sigma_r\nu \geqslant -2\epsilon \quad \text{for} \quad r = 4, 5, 9 \text{ and } 10. \tag{54}$$

Eliminating first ν and then κ from the first and last of the inequalities (54) we obtain

$$\left.\begin{aligned} (\rho_4|\sigma_{10}|+\rho_{10}\sigma_4)\kappa &\geqslant -2\epsilon\{|\sigma_{10}|+\sigma_4\}, \\ (\sigma_4|\rho_{10}|+\sigma_{10}\rho_4)\nu &\geqslant -2\epsilon\{|\rho_{10}|+\rho_4\}. \end{aligned}\right\} \tag{55}$$

Since
$$\begin{aligned} \rho_4|\sigma_{10}|+\rho_{10}\sigma_4 &= -\{\sigma_4|\rho_{10}|+\sigma_{10}\rho_4\} \\ &= \rho_4|\sigma_{10}|-\sigma_4|\rho_{10}| \\ &= \frac{(9-3\sqrt{7})\,3(24+9\sqrt{7})}{135+51\sqrt{7}} - \frac{(9+3\sqrt{7})\,3(24-9\sqrt{7})}{135-51\sqrt{7}} \\ &= \frac{27(3-\sqrt{7})(8+3\sqrt{7})}{3(3+\sqrt{7})(8+3\sqrt{7})} - \frac{27(3+\sqrt{7})(8-3\sqrt{7})}{3(3-\sqrt{7})(8-3\sqrt{7})} \\ &= \frac{9(16-6\sqrt{7})}{2} - \frac{9(16+6\sqrt{7})}{2} \\ &= -54\sqrt{7} < 0, \end{aligned}$$

the inequalities (55) show that
$$\kappa \leqslant C_5\epsilon, \quad \nu \geqslant -C_6\epsilon,$$

for some absolute constants C_5 and C_6. Substituting these bounds for κ and ν in the second or third of the inequalities (54) we obtain $\nu \leqslant C_7\epsilon$ and $\kappa \geqslant -C_8\epsilon$ for some constants C_7 and C_8. It now follows from (52) that (51) is satisfied for a suitable absolute constant C'.

LEMMA 9. *If* $\alpha > 3{\cdot}2$, *then* $\quad \alpha = \alpha_1, \quad \beta = \alpha_1'.$ (56)

Proof. We define α^* and β^* by the equations

$$\frac{1}{\alpha^*} = \frac{-1}{\alpha} - 2 + \sqrt{7}, \quad \frac{1}{\beta^*} = \frac{-1}{\beta} - 2 - \sqrt{7}. \tag{57}$$

Then, since
$$\frac{1}{\alpha_0}=\frac{-1}{\alpha_1}-2+\sqrt{7},\quad \frac{1}{\alpha_0'}=\frac{-1}{\alpha_1'}-2-\sqrt{7},\tag{58}$$

it follows from (51) that
$$|\alpha^*-\alpha_0|<C''\epsilon,\quad |\beta^*-\alpha_0'|<C''\epsilon,\tag{59}$$

for some absolute constant C''. Further, since $|\alpha_0\alpha_0'| = |\alpha_1\alpha_1'|$, it follows from (51) and (59) that
$$|\alpha^*\beta^*|\geqslant|\alpha\beta|(1-C'''\epsilon),\tag{60}$$

for some absolute constant C'''. Now, using (57), (10) and (60), we have for every integer ξ of $k(\sqrt{7})$,
$$\begin{aligned}|(\alpha^*\xi-1)(\beta^*\xi'-1)| &= |\alpha^*\beta^*|\left|\left(\xi-\frac{1}{\alpha^*}\right)\left(\xi'-\frac{1}{\beta^*}\right)\right|\\ &= |\alpha^*\beta^*|\left|\left(\xi+2-\sqrt{7}+\frac{1}{\alpha}\right)\left(\xi'+2+\sqrt{7}+\frac{1}{\beta}\right)\right|\\ &\geqslant\frac{|\alpha^*\beta^*|}{|\alpha\beta|}(1-\epsilon)\geqslant 1-\{1+C'''\}\epsilon.\end{aligned}\tag{61}$$

By the argument of lemma 5 it is clear that (59) and (61) imply that $\alpha^* = \alpha_0$ and $\beta^* = \alpha_0'$, provided ϵ_0 was chosen to be sufficiently small. Consequently, by (57) and (58), we must have $\alpha = \alpha_1$ and $\beta = \alpha_1'$.

5. LEMMA 10. *When a and b are of the form* (4), *there are no integers ξ of $k(\sqrt{7})$ for which*
$$|(\xi-a)(\xi'-b)|<\tfrac{9}{14},$$
but there are infinitely many integers ξ of $k(\sqrt{7})$ for which
$$|(\xi-a)(\xi'-b)|=\tfrac{9}{14}.$$

Proof. It clearly suffices to prove the result in the case when
$$a=\frac{1}{\alpha_0},\quad b=\frac{1}{\alpha_0'}.$$
We have then to prove that
$$\left|N\left(\xi-\frac{1}{\alpha_0}\right)\right|=\left|\left(\xi-\frac{1}{\alpha_0}\right)\left(\xi'-\frac{1}{\alpha_0'}\right)\right|\geqslant\frac{9}{14}$$

for all integers ξ of $k(\sqrt{7})$ and that this is satisfied with equality for an infinity of integers of $k(\sqrt{7})$. Since $N(\alpha_0) = \frac{14}{9}$, it is sufficient to prove that
$$|N(\alpha_0\xi-1)|\geqslant 1\tag{62}$$
for all integers ξ of $k(\sqrt{7})$ and that
$$|N(\alpha_0\xi-1)|=1\tag{63}$$
for an infinity of integers of $k(\sqrt{7})$. We have already seen in the proof of lemma 6 that (63) is satisfied by the integers η_n, ζ_n of $k(\sqrt{7})$ for every rational integer n.

Now, writing $\xi = x+y\sqrt{7}$,

$$N(\alpha_0\xi-1) = N\left(\frac{2\sqrt{7}}{5-\sqrt{7}}\xi-1\right)$$

$$= \frac{N(2\sqrt{7}\,\xi-5+\sqrt{7})}{N(5-\sqrt{7})}$$

$$= \tfrac{1}{18}N((14y-5)+(2x+1)\sqrt{7})$$

$$= \tfrac{1}{18}\{(14y-5)^2-7(2x+1)^2\} \qquad (64)$$

$$= \tfrac{1}{18}\{28(7y^2-5y-x^2-x)+18\}. \qquad (65)$$

But for any rational integers x, y

$$28(7y^2-5y-x^2-x)+18$$

is numerically greater than or equal to 18 unless it assumes the value -10. So (62) is satisfied for all integers ξ of $k(\sqrt{7})$ unless there are rational integers x, y satisfying

$$(14y-5)^2-7(2x+1)^2 = 28(7y^2-5y-x^2-x)+18 = -10. \qquad (66)$$

We now prove that there are no rational integers x, y satisfying (66). Suppose that x, y were such integers, and write $X = 2x+1$, $Y = 14y-5$. Then

$$Y^2 = 7X^2-10.$$

As X is odd, $\quad X \equiv \pm 1,\ \pm 3 \text{ or } 5 \pmod{10}.$

If $X \equiv \pm 1 \pmod{10}$, then

$$Y^2 \equiv 7-10 = -3 \pmod{10},$$

which is impossible. If $X \equiv \pm 3 \pmod{10}$, then

$$Y^2 \equiv 63-10 \equiv 3 \pmod{10},$$

which is impossible. Finally, if $X \equiv 5 \pmod{10}$,

$$Y^2 \equiv 0 \pmod{5},$$

and so Y is divisible by 5. This gives a contradiction since Y^2-7X^2 is divisible by 25 while 10 is not. These contradictions show that there are no rational integers x, y satisfying (66); and so the inequality (62) is satisfied for all integers ξ of $k(\sqrt{7})$. This proves the lemma and completes the proof of the theorem.

I am grateful to Professor Davenport for advice while I was working on this problem, and I am also indebted to Mr C. A. Rogers for his help in preparing the manuscript of this paper.

References

Davenport, H. 1946 *Proc. Acad. Sci. Amst.* **49**, 815–821.
Davenport, H. 1947 *Proc. Acad. Sci. Amst.* **50**, 378–389.
Heinhold, J. 1938 *Math. Z.* **44**, 659–688.
Heinhold, J. 1939 *Math. Z.* **45**, 176–184.
Varnavides, P. 1948*a* *Quart. J. Math.* **19**, 54–58.
Varnavides, P. 1948*b* *Proc. Acad. Sci. Amst.* **51**, 396–404, 470–481.
Varnavides, P. 1948*c* Thesis entitled 'Arithmetical properties of indefinite quadratic forms', approved for the degree of Ph.D. in the University of London, 1948.

Charge distribution and energy levels of trapped electrons in ionic solids

By J. H. Simpson,* *H. H. Wills Physical Laboratory, University of Bristol*

(*Communicated by N. F. Mott, F.R.S.—Received* 29 *October* 1948)

The charge distribution and energy levels of electrons trapped at two types of lattice defect in ionic solids (interstitial ions and *F*-centres) are calculated using a variation method which takes account of the fact that the field acting on the electron near such a defect depends upon the wave function of the electron itself. The model used for the interstitial ion consists of a point charge in a uniform polarizable medium while that for the *F*-centre consists of a potential well whose dimensions are decided by the charges and induced dipoles on the surrounding ions. The 1*s*-2*p* optical excitation energies are determined for *F*-centres in typical alkali- and silver-halides. Calculations of thermal activation energies from the 2*p* state to the conduction band and of thermal dissociation energies (1*s* state to conduction band) for the silver halides indicate that *F*-centres should play a more important part than interstitial ions in photoconductive processes in such materials.

1. Introduction

The distribution of the energy levels of electrons trapped at lattice defects in ionic crystals has been discussed by various authors (see, for example, Mott & Gurney 1940, chapter III). Calculations of the energies of these levels have been made for the case in which the defect is a vacant negative ion lattice point (*F*-centre) by Tibbs (1939) using a potential energy function of the form shown dotted in figure 6. Similar calculations might be made for the case in which the defect is an interstitial positive ion, although the flat portion of figure 6 would not then occur. Mott & Gurney have pointed out, however (as discussed below), that, in both cases, the field acting upon the electron near the lattice defect depends upon the wave function of the electron itself, so that this wave function and the associated potential can only be calculated by a self-consistent method. It is the purpose of this paper to present the results of the application of such a method to the determination of the 1*s* and 2*p* levels of an electron in the field of a positive interstitial ion, and in an *F*-centre. The charge distribution of the trapped electron and hence the effective radius of the impurity centre is also obtained.

2. Interstitial positive ion

The calculation of the wave function and energy level of an electron trapped by an interstitial ion is simpler than that for an *F*-centre, and will, therefore, be considered first.

An interstitial ion occupies a position between lattice points and is considered to be squeezed in so that the surrounding ions are displaced slightly to make room for it. If, to simplify the model used in the calculation, we consider the crystal as a continuum of dielectric constant κ, we may then regard the interstitial ion as a point charge inserted in this continuum, and the problem is to determine the wave function of an electron held to this positive charge by a Coulomb force in a polarizable medium. The assumption of spherical symmetry inherent in such a model is, of course, not strictly correct, and the deviations therefrom will be greatest in the immediate neighbourhood of the positive charge. The radial wave functions obtained can however be used to provide rough approximation to the energy levels of the interstitial ion and to show in a general way how the electronic charge is spread through the crystal.

---

* Now with National Research Council, Ottawa, Canada.

The field at radius r acting upon the electron trapped by this interstitial ion has been discussed in detail by Mott & Gurney (1940, p. 85). The calculation of this field is complicated by the fact that part of the polarization of the medium (that due to the displacement of the ions) remains constant during the short time of an electronic transition. Thus, the field acting on the electron when it is at radius r consists of two parts:

(1) The field $\epsilon/\kappa_0 r^2$ due to the positive ion, where κ_0 is the dielectric constant of the medium when the ions are held in their normal positions—the high-frequency ($> 5 \times 10^{12}$ sec.$^{-1}$) dielectric constant; and ϵ is the electronic charge.

(2) The field due to that part of the polarization of the medium that remains constant when the electron moves. This low-frequency part of the polarization, which is due to motion of the ions, will be produced by the interstitial ion charge and the fraction of the electronic charge that is normally within radius r. The latter quantity is $-\epsilon p(r)$, where

$$p(r) = \int_0^r |\psi|^2 4\pi r^2 dr, \tag{1}$$

if ψ is the wave function of the electron. Hence, the field at radius r due to this part of the polarization is

$$-\frac{\epsilon}{r^2}[1-p(r)]\left(\frac{1}{\kappa_0}-\frac{1}{\kappa}\right),$$

where κ is the static dielectric constant which is, of course, determined by both the low- and high-frequency parts of the polarization. Thus, the field acting upon the electron at radius r is

$$\frac{\epsilon}{\kappa_0 r^2}-\frac{\epsilon}{r^2}[1-p(r)]\left(\frac{1}{\kappa_0}-\frac{1}{\kappa}\right)$$

or

$$F(r) = \frac{\epsilon}{\kappa r^2}[1-p(r)]+\frac{\epsilon p(r)}{\kappa_0 r^2}. \tag{2}$$

It is seen from (2) that the field approaches $\epsilon/\kappa r^2$ for very small radii and $\epsilon/\kappa_0 r^2$ for very large radii, as we should expect.

The potential $V(r)$ to be used in Schrödinger's equation is the potential of the field F and thus depends upon ψ. Since a further relation between V and ψ is given by Schrödinger's equation itself, the two can only be calculated by a self-consistent method. The actual method used was of the variation type and may be outlined briefly as follows:

(a) The form of the wave function of the electron is estimated and a function of this type having one or more variable parameters is chosen. In this case, since the arrangement is similar to that of the hydrogen atom, an initial function of the form

$$\psi_0 = A\,e^{-\mu r} \tag{3}$$

is chosen for the 1*s* state when one parameter is used. This wave function applies, strictly speaking, to the case of an interstitial hydrogen nucleus. The way in which the wave function of the interstitial ion varies near $r = 0$ is not known, but for a calculation of the type considered here a wave function of the hydrogen form should be sufficiently accurate. The constant A is evaluated in terms of μ by normalization.

(*b*) Using equations (1) and (2), the field in which the electron moves and hence its potential $V_\mu(r)$ may be calculated.

(*c*) We may now assume that the final wave function obtained from Schrödinger's equation is similar in form to ψ_0 and write it

$$\psi_1 = B\,e^{-\lambda r}. \tag{4}$$

Using the variation integral

$$W = \int \psi_1^* \left[-\frac{\hbar^2}{2m}\nabla^2 + V_\mu \right] \psi_1 \, d\tau,$$

an upper limit to the energy of the 1*s* state may be obtained. W is a function of λ and μ.

(*d*) The best function of form (4) is obtained by putting $\partial W/\partial\lambda = 0$. This gives λ as a function of μ.

(*e*) If our final solution of the wave equation is to be consistent with the function initially chosen, we must have $\lambda = \mu$, and hence from (*d*) we obtain a relation which can be solved for λ alone. Substitution in (4) and in the expression for W obtained from the variation integral gives the required wave function and energy.

2·1. *Detailed calculations—interstitial ion* 1*s state*

Applying the above method and using the wave function given by (3), we have

$$\psi_0 = A\,e^{-\mu r}, \quad A^2 = \mu^3/\pi, \quad p(r) = 1 - e^{-2\mu r}(2\mu^2 r^2 + 2\mu r + 1),$$

and, in atomic units (Hartree 1928),

$$F = \frac{1}{\kappa r^2}[1 - p(r)] + \frac{1}{\kappa_0 r^2} p(r)$$

$$= \frac{1}{\kappa_0 r^2} - \frac{1}{r^2}\left(\frac{1}{\kappa_0} - \frac{1}{\kappa}\right) e^{-2\mu r}(2\mu^2 r^2 + 2\mu r + 1).$$

The potential energy term to be used in the variation integral is, therefore,

$$V_\mu = \int_\infty^r F\,dr = -\frac{1}{\kappa_0 r} + \frac{1}{r}\left(\frac{1}{\kappa_0} - \frac{1}{\kappa}\right)(1 + r\mu)\,e^{-2\mu r}.$$

The final wave function is $\quad \psi_1 = B\,e^{-\lambda r}, \quad B^2 = \lambda^3/\pi.$

The energy is given by $\quad W = \int \psi_1[-\tfrac{1}{2}\nabla^2 + V_\mu]\,\psi_1\,d\tau.$

Since all functions involved are spherically symmetrical the volume element may be taken as $d\tau = 4\pi r^2 dr$. The radial form of $[-\tfrac{1}{2}\nabla^2 + V_\mu]\,\psi_1$ is

$$\left[-\frac{1}{2r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial \psi_1}{\partial r}\right) + V_\mu \psi_1 \right],$$

and, following integration, the value of W is found to be

$$W = \frac{\lambda^2}{2} - \frac{\lambda}{\kappa_0} + \frac{\lambda^3}{(\mu+\lambda)^2}\left(1+\frac{\mu}{\mu+\lambda}\right)\left(\frac{1}{\kappa_0}-\frac{1}{\kappa}\right).$$

Evaluating $\partial W/\partial\lambda$ and putting $\lambda = \mu$ in $\partial W/\partial\lambda = 0$, the following value for λ is obtained:

$$\lambda = \frac{1}{16}\left(\frac{5}{\kappa_0}+\frac{11}{\kappa}\right),$$

whence
$$W = -\frac{1}{512}\left(\frac{5}{\kappa_0}+\frac{11}{\kappa}\right)\left(\frac{15}{\kappa_0}+\frac{1}{\kappa}\right) \text{ atomic units}$$

$$= -0{\cdot}0530\left(\frac{5}{\kappa_0}+\frac{11}{\kappa}\right)\left(\frac{15}{\kappa_0}+\frac{1}{\kappa}\right) \text{ electron volts.}$$

Values of λ in atomic units and W in electron volts for some ionic crystals are given in table 1. Corresponding quantities for the hydrogen atom in empty space are included for comparison.

TABLE 1. CONSTANTS FOR INTERSTITIAL ION 1s STATE

substance	λ (in atomic units)	W (eV)	κ_0	κ	Bohr orbit radius $1/\lambda$	
					atomic units	(Å)
NaCl	0·261	−1·52	2·25	5·62	3·8	2·0
AgBr	0·120	−0·34	4·62	13·1	8·3	4·4
AgCl	0·134	−0·44	4·01	12·3	7·5	4·0
AgCl	two-parameter method	−0·44	4·01	12·3	7·4	3·9
H	1·0	−13·61	—	—	1·0	0·54

The radial distribution of charge $(4\pi r^2\psi^2)$ for NaCl and AgBr is plotted in figure 1. It is doubtful whether interstitial ions can exist in appreciable numbers in NaCl, but it is probable that they are important in silver halides (see Mott & Gurney 1940, chapter II; or Seitz 1940, chapter XV). The charge distribution for AgCl is shown, to a different scale, in figure 2. These results indicate that the effect of the polarizability of the medium is to increase the radius of the Bohr orbit by a factor approximately equal to $3/(\kappa_0^{-1}+2\kappa^{-1})$, and to decrease the ionization energy by the (approximate) factor $\kappa_0/0{\cdot}16(\kappa_0^{-1}+2\kappa^{-1})$. The Bohr orbit of the 1s state, therefore, has a radius of $1{\cdot}6/(\kappa_0^{-1}+2\kappa^{-1})$ Å, and its energy is $-4{\cdot}35(\kappa_0^{-1}+2\kappa^{-1})/\kappa_0$ electron volts.

2·2. *Interstitial ion* 1s *state—two-parameter solution*

The variation method gives an upper limit to the energy but does not indicate how far this is from the true value. Some indication of the accuracy of the values in table 1 can, however, be obtained by comparison with the solution when the initially chosen wave function has two parameters. If the results lie reasonably close together we may assume that they give a fair approximation to the actual energy value. A two-parameter calculation was, therefore, carried out. This was considerably more

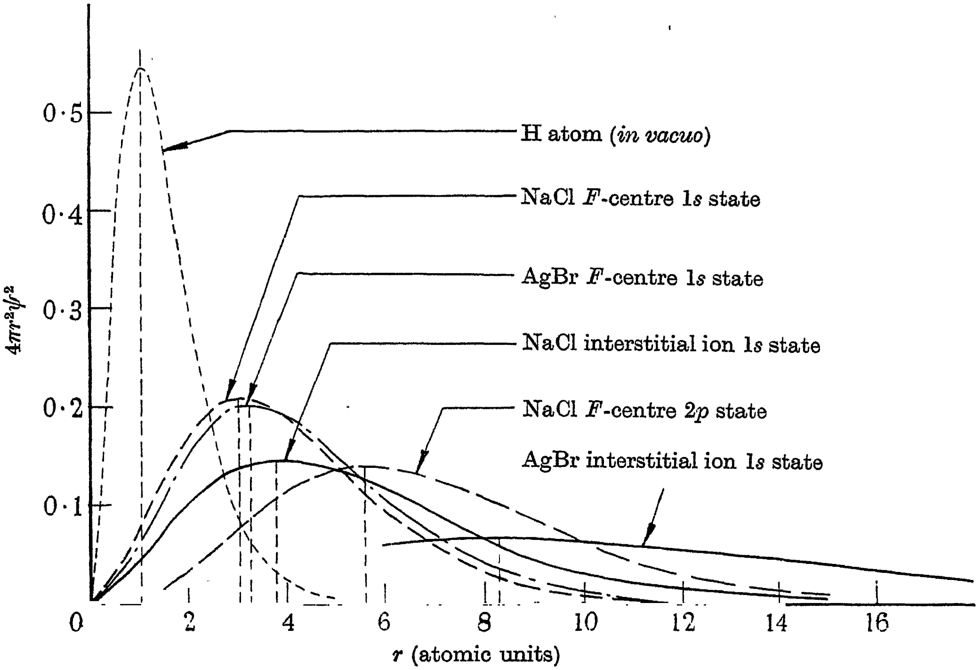


FIGURE 1. Charge distribution of electrons trapped by interstitial ions and *F*-centres (dashed vertical lines indicate Bohr radii).

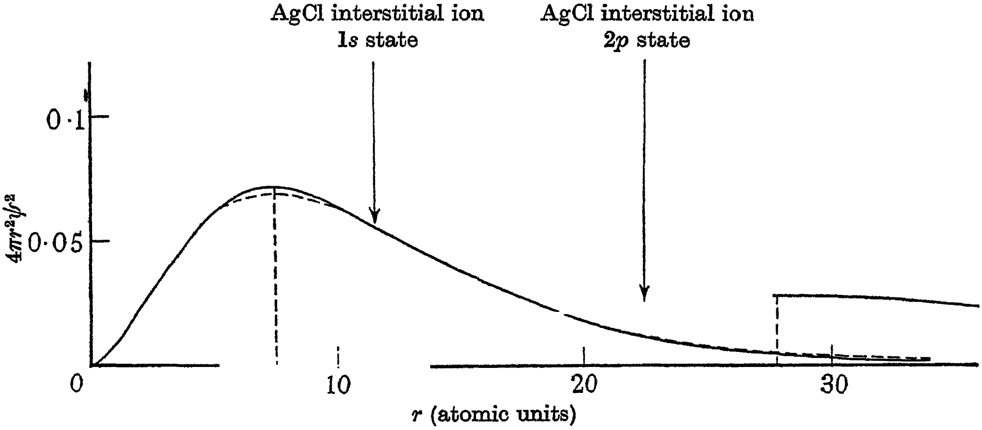


FIGURE 2. Charge distribution of an electron trapped by an interstitial ion. Full line: one-parameter solution. Dashed lines: two-parameter solution. Dashed vertical lines indicate Bohr radii.

laborious than that for the one-parameter case, and the values of the parameters could only be determined by numerical methods. This involved substituting the values of κ_0 and κ early in the calculation, and results were obtained for one substance only, namely, AgCl.

The wave function chosen had the form

$$\psi = Ar^{-1}(e^{-\alpha r} - e^{-\beta r}),$$

and the values of the parameters in atomic units were

$$\alpha = 0{\cdot}116 \quad \text{and} \quad \beta = 0{\cdot}154.$$

The energy was found, within the accuracy of the calculation, to be equal to that obtained by the one-parameter method, and it, therefore, seems reasonable to assume that the values in table 1 are substantially correct. The distributions of charge obtained by the two methods are compared in figure 2.

2·3. *Interstitial ion, 2p state*

The energy and wave function for the $2p$ state have been determined using the same model, and again considering the electric field to be spherically symmetrical. The wave functions chosen contain one parameter and are again the analogues of the hydrogen atom wave functions. Thus

$$\psi_{2z} = Ar\,\mathrm{e}^{-\alpha r}\cos\theta, \quad \psi_{2x} = Ar\,\mathrm{e}^{-\alpha r}\sin\theta\,\cos\phi, \quad \psi_{2y} = Ar\,\mathrm{e}^{-\alpha r}\sin\theta\,\sin\phi.$$

The calculation is similar to that for the $1s$ state, although somewhat more complicated. The values of α and the energies are given in table 2, and the equivalent spherically symmetrical electronic charge distribution for AgCl is shown in figure 2.

TABLE 2. CONSTANTS FOR INTERSTITIAL ION $2p$ STATE

substance	α (atomic units)	W (eV)	mean Bohr orbit radius $2/\alpha$ atomic units	(Å)
NaCl	0·137	−0·44	14·6	7·7
AgBr	0·063	−0·10	31·7	16·8
AgCl	0·072	−0·13	27·8	14·8

The electronic charge distribution is, of course, spread over many more lattice distances in this case. For the silver halides it is still appreciable at distances equal to ten times the distance between nearest neighbours. The $2p$ level for these substances is separated from the conduction band by less than 0·2 eV.

3. *F*-CENTRE

An F-centre consists of an electron trapped at a point in the lattice from which a negative ion is missing. In this case the calculation of the potential energy of an electron in the neighbourhood of the trapping point is more complicated, but we can obtain an idea of its general form by considering in detail the effect of the immediately adjacent ionic charges. Thus in figure 3 let O be the site of the missing negative ion and A, B, C, D, E, F those of the adjacent positive ions. Then, considering for the moment that the latter act as point charges, the potential at point P a distance r from O is

$$\phi = \frac{\epsilon}{a+r} + \frac{\epsilon}{a-r} + \frac{4\epsilon}{\sqrt{(a^2+r^2)}},$$

where a is the distance between nearest neighbours in the lattice. The potential energy of an electron at P is, therefore (in atomic units),

$$V = \frac{1}{a}\left\{\frac{2}{(1-r^2/a^2)} + \frac{4}{\sqrt{(1+r^2/a^2)}}\right\}.$$

The potential energy along line OB is plotted in figure 4 for $0 \leqslant r \leqslant a$. It is apparent that, except for the singularity in potential at point B ($r = a$), the function is relatively flat over practically the whole range.

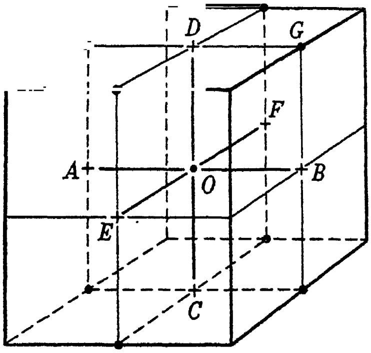


FIGURE 3. Ions surrounding a vacant lattice point. Nearest neighbours (positive ions) represented by plus signs; next nearest neighbours (negative ions) by dots.

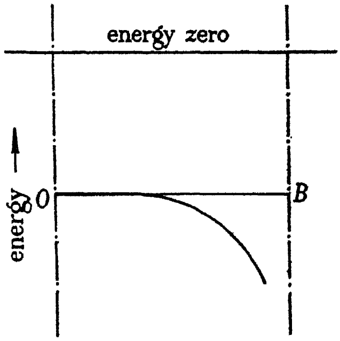


FIGURE 4. Potential energy of an electron along line OB of figure 3.

The effect of the more distant shells of ions could be included in a similar manner. The potential produced by these in the range $0 \leqslant r \leqslant a$ will vary even less than that in figure 4.

We must also consider the effect of the dipoles induced on the ions surrounding O because of the effective charge at O when the ion there is removed. The moments of these dipoles for NaCl may be obtained from the results of Mott & Littleton (1938). Those of the first shell of (positive) ions are very small. They produce a potential at O equal to 0·5 % ϕ_1, where ϕ_1 is the potential produced by the charge on the ions. The potential at O due to the dipoles of the second shell of (negative) ions is larger (4·5 % ϕ_1). It can be shown quite readily, however, that, except for the singularity at $r = a$, the curve of potential due to both these sets of dipoles is flat within 20 %

for distances from 0 up to 0·9a. Succeeding shells of dipoles will, of course, produce potential curves that are still flatter.

The above discussion applies to points on the three axes of figure 3. Similar calculations can be made for other parts of the region surrounding O, however, and give substantially the same result. The general form of the potential-energy function for two different crystallographic directions is shown in figure 5. The smoothed-over potential function is therefore that given by the dashed lines, the average radius of the flat base being approximately equal to a.

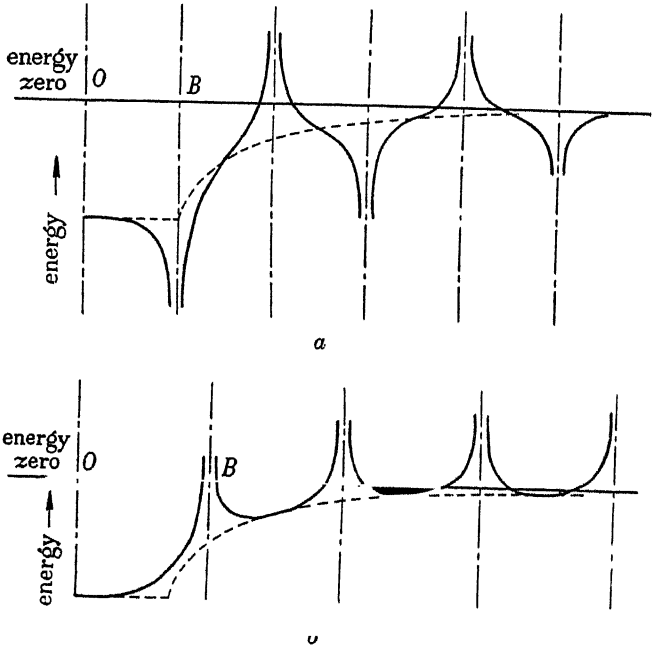


Figure 5. Potential energy of an electron along two different crystallographic directions. a, potential energy along line OB of figure 3. b, potential energy along line OG of figure 3.

In considering the effect of the surrounding medium on the electronic charge distribution it is convenient to use Jost's well-known model in which the material is treated as a uniform medium containing a spherical hole at the position of the missing ion. The value of the potential at the centre of this hole is then given by (see Tibbs 1939)

$$\phi = \frac{\alpha\epsilon}{a} - \frac{\epsilon}{R}\left(1 - \frac{1}{\kappa_0}\right),$$

where $\alpha = 1{\cdot}746$ is Madelung's constant and R is the radius of the hole. The potential energy of an electron at this point is, therefore, in atomic units,

$$V = -\frac{\alpha}{a} + \frac{1}{R}\left(1 - \frac{1}{\kappa_0}\right).$$

The second term of this is the energy gained from the polarization of the medium by the effective charge at the centre of the sphere when the electron is removed. The high-frequency dielectric constant κ_0 is used, since we are considering optical transitions, during which the ions have not time to move.

The value of V and hence the radius of the sphere may be obtained from the results of Mott & Littleton (1938). For NaCl the values are $V = -0{\cdot}22$ atomic unit and $R = 0{\cdot}95a = 5{\cdot}0$ atomic units. The radius of the flat portion of the potential curve discussed above is of the order of R, and for the purpose of this paper it seems reasonable to assume that they are equal. The picture of the potential energy curve is, therefore, as shown dashed in figure 6. For large distances from the origin it will have the form $-(\kappa_0 r)^{-1}$. For distances just greater than R its form will be difficult to determine, and if account is taken of the effect of the electron charge distribution on the dielectric medium in the same manner as for the interstitial ion, a discontinuity will be introduced at $r = R$, since the methods of calculating potential in the regions $0 \leqslant r \leqslant R$ and $R \leqslant r \leqslant \infty$ bear no relation to each other. The curve obtained will be that shown in the full line in figure 6, whereas the true curve is probably more like that shown dashed. Some error will be introduced into the calculation because of this difference but we should expect it to be relatively small.

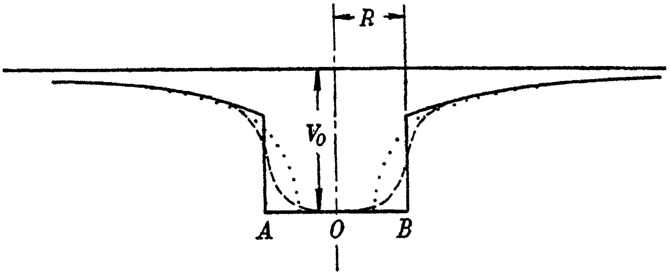


FIGURE 6. Potential energy functions of an electron in an F-centre. Full line: potential function used in this paper. Dashed line: probable form of true smoothed-over potential energy function. Dotted line: potential-energy function used by Tibbs.

For the $2p$ state a much more serious error results from the assumption that the field of the trapping centre is spherically symmetrical and from the smoothing-over of the potential functions of the first shell of (positive) ions. It is to be expected that these ions would have a pronounced effect upon the shape of the $2p$ wave function, causing it to move outwards from the trapping centre and to spread over a large volume of the crystal than in the case of the simple model described above. As a result, the $2p$ energy level calculated using that model will be too low, i.e. the magnitude of the energy value obtained will be too large.

Tibbs has obtained a potential curve for the F-centre using a curve of Coulomb form $\{-(\kappa_0 r)^{-1}\}$ and determining the radius of the bottom of the potential energy well from the intersection of this curve with the line AB. In view of the above estimate of the potential surrounding O, however, it seems that the form of function given by the full line should be more satisfactory.

3·1. *Detailed calculations, F-centre* 1*s and* 2*p states*

Using the same functions as in the one-parameter interstitial ion case the value of W was determined for the $1s$ and $2p$ states of F-centres in NaCl and AgBr. The variation integral now consists of three parts:

$$W = \int_0^\infty \psi_1(-\tfrac{1}{2}\nabla^2\psi_1)\,4\pi r^2\,dr + \int_0^R \psi_1 V_0 \psi_1\,4\pi r^2\,dr + \int_R^\infty \psi_1 V_\mu \psi_1\,4\pi r^2\,dr,$$

where the function V_μ has the same form as in the previous case. Values for W and λ were obtained as before, except that, because of the complicated form of the variation integral, it was necessary to solve the equations numerically.

To test these results by two-parameter calculations would involve a prohibitive amount of labour. It is possible, however, to use a more accurate one-parameter wave function for the 1*s* state. For a potential function of the form used we should expect that $d\psi/dr = 0$ at $r = 0$, and we may choose an initial wave function of the form $\psi_2 = A\,e^{-\alpha r}(1+\alpha r)$ which satisfies this condition. The resulting increase in the magnitude of W is about 7 % for both NaCl and AgBr, so that it is probably safe to assume that these new values are approximately correct. For the 2*p* state a wave function of the form used in the interstitial ion case should be quite satisfactory. The values obtained for W and λ are given in table 3. Data on the 1*s* state refer to the more accurate wave function.

TABLE 3. CONSTANTS FOR *F*-CENTRE 1*s* AND 2*p* STATES

	λ (atomic units)		W (eV)		mean Bohr orbit radius (atomic units)	
substance	1*s*	2*p*	1*s*	2*p*	1*s*	2*p*
NaCl	0·52	0·36	−3·2	−1·0	3·0	5·6
AgBr	0·49	0·30	−2·1	−0·26	3·3	6·7

Results for AgCl will be similar to those for AgBr.

The charge distributions for the 1*s* and 2*p* states of the *F*-centre are compared with those of the interstitial ion in figure 1. The following points may be noted:

(*a*) As pointed out above the magnitude of W_{2p} is probably considerably larger than it should be.

(*b*) The distribution of charge of an *F*-centre is confined to a smaller volume of the crystal than that of an interstitial ion.

(*c*) For the *F*-centre 1*s* state the amount of charge lying outside of $r = R$ is relatively small, and the effect of the low-frequency part of the polarization of the medium will likewise be small. For the model used it would probably be a sufficiently accurate approximation to assume that the effective dielectric constant of the medium is κ_0.

4. EXCITATION AND ACTIVATION ENERGIES

It is of interest to determine the energy of the 1*s* to 2*p* optical transition for an *F*-centre in NaCl using the above model. Since during the short time of an optical transition the surrounding ions have not time to move, it is necessary in this case to obtain the energy of that state in which the electron has a 2*p* wave function but is acted upon by the field corresponding to a 1*s* wave function. We may call this energy $W_{2p(1s)}$ and the required energy of the optical transition is $(W_{2p(1s)} - W_{1s})$, where W_{1s} is the energy of the ground state. The values for W_{1s} and $W_{2p(1s)}$ for NaCl are respectively $-3{\cdot}2$ and $-1{\cdot}0$ eV, so that the energy of the $1s$ to $2p$ optical transition is 2·2 eV. For reasons given above we should expect this to be a lower limit. The experimentally determined value for this quantity is 2·7 eV. In view of the simplicity of the model used and the approximations made, the agreement between theoretical and experimental values may be considered satisfactory. Results for KCl give somewhat better agreement with the experimental value.

After the electron has made the transition the ions surrounding the trapping centre will no longer be in equilibrium. The new equilibrium arrangement will be the one which corresponds to the $2p$ wave function of the electron and the field acting upon the electron, and its potential energy will, therefore, be changed. The electron energy level will, therefore, move slightly closer to the condition band when the ions take up their new positions. For the $1s$ to $2p$ transition in an F-centre the change of energy thus occurring is very small, since the value of $W_{2p(1s)}$ is, within the accuracy of the calculation, equal to W_{2p}, the energy corresponding to the final $2p$ equilibrium state. This is due to the fact that the wave function and potential are both small at $r > R$, so that the effect of the medium upon the value of the variation integral is relatively small whether the field corresponds to a $1s$ or a $2p$ wave function.

A trapped electron in an excited state may receive sufficient energy from the lattice vibrations to put it into the conduction band. For such a thermal process the ions surrounding the impurity centre will be in equilibrium before and after the act, and the energy required will be less than if the transition were an optical one. This is due to the fact that in moving to new equilibrium positions the ions give out energy. The amount of this energy may be calculated by a modification of a method described by Mott & Gurney (1940, p. 161). The energy of an element of the medium of radius r after the electron has been moved but before the surrounding ions take up new positions is

$$\frac{1}{8\pi}ED\,4\pi r^2\,dr = \frac{r^2\,dr}{2}\left\{\frac{\epsilon}{\kappa r^2}+\frac{\epsilon p(r)}{r^2}\left(\frac{1}{\kappa_0}-\frac{1}{\kappa}\right)\right\}\frac{\epsilon}{r^2},$$

where the field E (in curly brackets) is given, using the value of $p(r)$ for the $2p$ state, by equation (2), and the displacement $D = \epsilon/r^2$. The energy of the same element after the ions have taken up their new positions is

$$\frac{r^2\,dr}{2}\,\frac{\epsilon}{\kappa r^2}\,\frac{\epsilon}{r^2}=\frac{dr}{r^2}\,\frac{\epsilon^2}{2\kappa}.$$

Hence the energy given up by the medium is, in atomic units,

$$U_1=\int_0^\infty \frac{1}{2}\left[\frac{1-p(r)}{\kappa}+\frac{p(r)}{\kappa_0}-\frac{1}{\kappa}\right]\frac{dr}{r^2}=\int_0^\infty \frac{1}{2}\left(\frac{1}{\kappa_0}-\frac{1}{\kappa}\right)\frac{p(r)\,dr}{r^2},$$

when the trapping centre is an interstitial ion, and

$$U_2=\int_R^\infty \frac{1}{2}\left(\frac{1}{\kappa_0}-\frac{1}{\kappa}\right)\frac{p(r)\,dr}{r^2},$$

when it is an F-centre. Using the formula

$$E = -W_{2p} - U,$$

where E is the thermal activation energy, the results given in table 4 were obtained.

TABLE 4. THERMAL ACTIVATION ENERGY ($2p$ STATE TO CONDUCTION BAND)

type of centre	energy released by medium U (eV)	thermal activation energy E (eV)
NaCl, F-centre	0·55	0·45
AgBr, interstitial ion	0·06	0·04
AgBr, F-centre	0·25	0·01

It thus appears that an electron trapped by an interstitial ion or F-centre in AgBr or AgCl has a high probability of reaching the conduction band once it has been excited. For NaCl thermal activation from the excited state to the conduction band appears to be less probable. It is to be emphasized, however, that the F-centre figures in the final column of table 4 should be accepted with reserve, since the energy calculated for the $2p$ state may be in error. Mott & Gurney give evidence that, for the NaCl F-centre, the thermal activation energy is of the order of 0·1 eV, indicating that the value in table 4 is too high, as expected.

5. TYPES OF TRAPPING CENTRES IN SILVER HALIDES

The calculations of this paper give interesting information concerning the relative importance of interstitial ions and F-centres in silver halides. The energy of the $1s$ state for an electron trapped by an interstitial ion is −0·4 eV. This is the energy which would be required to dissociate the trapping centre optically. To determine whether an electron will remain trapped by an interstitial ion for any length of time we must obtain the thermal dissociation energy, i.e. we must take account of the energy gained because of polarization of the medium when the electron is removed. When this is done it is found that for both AgBr and AgCl the thermal dissociation energy is approximately 0·1 eV. Thus at ordinary temperatures most interstitial ions will have no electrons trapped in their fields. For F-centres, on the other hand, the optical dissociation energy is about 2 eV and the thermal dissociation energy 1·7 eV, so that very few F-centres will be dissociated. It thus appears that, other factors (such as the ease of formation of defects) being equal, F-centres will play a more important part than interstitial ions in photoconductive processes in silver halides.

In conclusion, it may be observed that the results of this paper are in moderately good agreement with available experimental data, and that for those cases in which comparison with experiment could not be made the results are in quite reasonable agreement with theoretical ideas on the subject. It may be expected, therefore, that the squares of the wave functions obtained give a fair indication of the way in which the electronic charge is spread through the crystal.

The writer wishes to express his gratitude for the interest and guidance of Professor N. F. Mott who suggested this problem. Acknowledgement is also made to the National Research Council of Canada for the grant which enabled the writer to undertake the work.

REFERENCES

Hartree, D. R. 1928 *Proc. Camb. Phil. Soc.* **24**, 105.
Mott, N. F. & Gurney, R. W. 1940 *Electronic processes in ionic crystals.* Oxford: The Clarendon Press.
Mott, N. F. & Littleton, M. J. 1938 *Trans. Faraday Soc.* **34**, 485.
Seitz, F. 1940 *The modern theory of solids.* New York: McGraw-Hill Book Company, Inc.
Tibbs, S. R. 1939 *Trans. Faraday Soc.* **35**, 1471.

Spontaneous deformation of protocatechuic acid crystals

By R. W. Wood, For. Mem. R.S.

(*Received* 21 *October* 1948)

[Plates 7 to 12]

In this paper an extensive study is reported of the very remarkable, and thus far apparently unique, case of the deformation in three dimensions of protocatechuic acid, to which attention was drawn many years ago by Otto Lehmann.

The deformations are spontaneous, and are probably due to progressive gliding of the lattice planes, which exist in two configurations, one stable and the other unstable, the latter being the condition of the long prismatic rods when they first form.

Such a prism presently deforms into a zigzag crystal, with stable and unstable sections in alternation which, with continuation of the deformation, becomes again straight, but now in the stable configuration. The bending is progressive, like that of an umbrella case, pendant from the end of an oblique cane pointed down, when the latter is pushed into it. The movements are so rapid that motion pictures, made with a microscope, were necessary for the observation of certain stages of the deformation. The deformations have been shown to many chemists and physicists during the past decade or more, none of whom had ever seen or heard of this remarkable type of crystal movement. The deformations are usually observed as the warm saturated solution cools, but they also occur after the crystal has been dried for many hours.

Introduction

The very remarkable and I believe unique phenomenon of the spontaneous deformation of crystals of protocatechuic acid, (3,-4-dihydroxybenzoic acid), forming in a warm saturated solution in water as it cools, was very briefly described by Lehmann in Frick (1890, 1905). I assume that Lehmann was the first to observe and describe the behaviour of these crystals but have been unable to find any mention of the phenomenon elsewhere, and he makes no reference in Frick's book to any earlier paper on the subject.

Lehmann placed a small amount of the powder, together with one or two drops of water, on a microscope slide, lowering a watch-glass, convex side down, on to the mixture, which was then warmed until the crystals dissolved. On cooling, minute rod-like prisms formed, which presently underwent spontaneous deformations, illustrated by his diagrams reproduced in figure 1, which, however, completely fail to represent what actually takes place, giving the impression of a sudden bend at a fixed point.

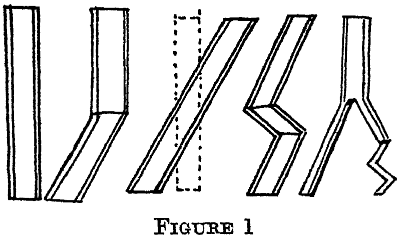


Figure 1

Equally misleading was his description of the phenomenon 'Die Verschiebung erfolgt so rasch und leicht, dass man glauben könnte, eine flüssige Masse vor sich

Linlithgow Library.

zu haben, durch welche sich momentan eine Erschütterung fortplanzt'. He thus appears to have classified them with the so-called 'fluid crystals', with which he was much occupied at the time, whereas they are quite as solid as the crystals of most organic compounds, and the actual movement producing the bending of crystal 3, shown by motion pictures in figures 16 to 19, plate 8, is similar to that of a pendant umbrella case when a cane, inclined at an angle of 30° with the vertical, is pushed down into it.

My attention was drawn to the curious movements of these crystals by Professor H. Rubens, when I was in Berlin in 1910, and I was given a small amount of the material, prepared, I think, by Kahlbaum. My first observations were made with a low-power microscope, employing Lehmann's technique. The most spectacular stage of the recrystallization occurred several minutes after the growths had started, slender, jointed, prismatic rods growing off obliquely from a thick stalk, bending and straightening like the claws of an active shrimp, so rapidly that the eye could scarcely follow the details of the movements. For two decades or more I showed this extraordinary performance to a large number of physicists and chemists, none of whom had ever seen or heard of the phenomenon before.

Having nearly exhausted the material obtained in Berlin, motion pictures of the formation and deformation of the crystals were made with a microscope fitted with a 32 mm. cinematograph camera. These I showed at a meeting of the National Academy of Science and at the Cavendish Laboratory in Cambridge in 1938, but published nothing about them at the time.

Types of crystals formed

Several different types of crystals are usually formed, depending on the saturation of the solution, temperature to which it has been heated, and other factors. Triclinic, isolated crystals, resembling rhombs of Iceland spar, which, for brevity, will be referred to as rhombs, are shown in figures 23 to 27, plate 9; also the jointed rods formed from unstable straight rods by the development of stable segments. These are usually based on rhombs as shown in figures 25 and 26 or found sprouting obliquely, as claws from a thicker crystalline stalk, of the stable type, usually more or less cylindrical, with a smooth surface, shown in figures 11 to 22, plate 8. The moving claws result from deformations of the oblique rods, and eventually become parallel to the main stalk, crowding as close to it as possible. These transitions will be discussed more fully later.

Motion pictures of the deformations

We will now consider the details of the motion pictures made with the material obtained in Berlin, which I have never succeeded in duplicating exactly with the more recently obtained acid from different sources, the thick cylindrical main stalks with smooth surfaces never having appeared again.

Figure 3, plate 7, shows the stable main stalk, with the unstable crystal growing out from it at an angle of about 30° like an ear of corn, and marked 1. In figure 4 a short kink is seen, about half-way down, where the transformation into the stable modification is taking place; this portion is seen to be parallel to the main stalk, and its length increases in figures 5 and 6, and in figure 7 the entire upper portion is stable. No downward movement of the deformation is noticeable up to figure 6, but in figure 7 it has started, and has carried the overlying portion to the left, bringing its peak into contact with a smaller unstable crystal. With continued downward movement of the deformation the peak presses strongly against the small crystal fracturing it in figure 8, and crowding into the closest possible contact with the main stalk in figure 9. The next picture shows crystal 1 deflected back, and nearly in its original position. To account for this it will be necessary to formulate a theory explaining the movements which take place among the molecules of the crystal during its deformations.

THEORY OF THE DEFORMATIONS

It seems probable that the simplest explanation of the phenomena is to assume that parallel superposed gliding planes of the crystal lattice may be stationed in two different positions with respect to each other, one stable and the other unstable. The diagram, figure 2, represents the positions of the planes for the condition shown by the unstable crystal 1 in figure 4, the black circles representing the molecules of the acid. The main stalk is represented by the vertical column while the oblique unstable crystal is made up of three parts, two oblique unstable portions, separated by a vertical stable one, the former showing a suggestion of internal hexagonal structure, the latter cubical.

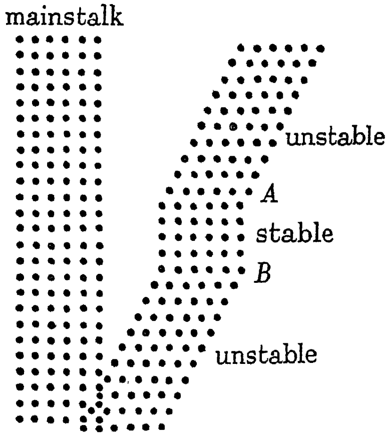


FIGURE 2

A shift to the left of the plane *A*, through a distance of one-half of the molecular diameter, will bring the plane into the stable position, and in making the movement it will carry the portion of the crystal above it to the left through an equal distance. The repetition of this process in succession by the overlying planes will bring the entire upper part of the crystal into the vertical and stable position, parallel to the main stalk. Plane *B* then shifts to the left, carrying everything above it in the same direction, this process proceeding downward and bringing the now stable crystal into contact with the main stalk.

Very often, under proper illumination, one can see a narrow dark line separating the stable from the unstable portion, and in the case considered the dark lines move in opposite directions, one up, the other down. Obviously the trapping of an unstable segment between two stable ones can also occur if two well-separated lattice planes of the unstable crystal begin to shift simultaneously the successive shifts to the left, proceeding downward from the upper part of the crystal, approach a similar series of shifts from a lower part of the crystal moving upward, the dark lines which separate the joints approaching, instead of receding, from each other, as in the former case.

Events following the embedding of an unstable crystal in a stable one

To account for the return of the deflected crystal 1 to its original position, as shown in figure 10, plate 7, and mentioned earlier in the paper, we must bear in mind that both the main stalk and the unstable obliqué crystal 1 based on it, have been growing for some time before the deformations had commenced. This results in the embedding of the lower part of the unstable crystal in the stable surface layers of the main stalk, and it is in this region that the shifting planes in the unstable crystal meet a resistance that results in a catastrophe, ruptures of the lattice occurring that have thrown the crystal back into the position shown in figure 10. This must have occurred in about $\frac{1}{20}$ sec., as figures 9 and 10 were adjacent 'frames' on the film.

This process of lattice rupture is apparently shown by photographs made years later (figures 25 and 26 of plate 9), showing an unstable rod growing from a rhomb. On the original photograph the embedded portion of the rod is seen through the perfectly transparent rhomb, while in a second picture, made 10 min. later, the lower part of the rod, which was originally nearly parallel to the long diagonal of the 'diamond', has passed over into the stable form, parallel to an edge of the diamond, the portion embedded in the rhomb pulverizing the crystal locally, and causing a black spot opaque to light.

A second sequence, in which the events occur in reverse order from those shown on plate 7, is reproduced on plate 8. Here the lattice rupture is the first cause of the bending of the unstable crystal 3 to one side. We are to follow the movements of three crystals 1, 2 and 3, the latter being the one causing lattice rupture. Numbers 1 and 2 are side by side, number 2 a little behind number 1. On this plate, figures 11, 12, 13, and 17, 18, 19, are slightly out of focus, and show the crystals as transparent bordered by black lines, figures 14, 15, and 16 are in focus and show the crystals as opaque, with distinct surface markings.

In figure 12, rod 2 is passing behind rod 1 and in figure 13 has become stable and parallel to the main stalk, visible in part through rod 1, slightly displaced by refraction. While this has been going on, the oblique portion of the main stalk *AB* in figure 11, evidently still unstable, has passed over in figure 12 to the stable condition, and is now parallel to the stable portion of the stalk below *A*, which process has carried 1 and 2 to the left, leaving a gap between 2 and 3. It is probable that the part *AB* was formed by the fusion of the unstable oblique crystals 1 and 2. In figure 13 this gap is much wider, the stable condition having ascended to the point indicated by an arrow, while in figure 14 it has nearly reached the upper edge of the figure. In figure 14 the deformation of unstable rod 3 has commenced, but in this case the shift of the lattice planes originated at the bottom of the rod, where it was embedded in the surface growth of stable stalk, and ruptures occurred almost at the moment at which the picture was taken, throwing rod 3 over to the right as shown in figure 15, an event similar to that in the previous sequence in figure 10, plate 7. In figure 14 the only apparent change is in the surface markings at the point where the rod joins the main stalk, the loss of transparency at the tip of the rod being due to improved focus as mentioned above. In figure 15 the lower portion of the rod has become stable, and the further rise of the line of demarcation between the stable and unstable parts is shown in the following figures, 'out of focus' transparency recurring in figure 17 and persisting up to figure 19, where the entire rod has become stable and straight; figure 20, taken several minutes later, shows the rod opaque, and sprouting, oblique, unstable finger-like crystals which I shall call 'dactyls'. Series 15 to 19 took place in about 7 sec. Figures 20, 21 and 22, added after a recent examination of the 32 mm. film, show the further developments. In figure 22 one of the dactyls has developed into an unstable crystal marked *U*, the other two 'deformed' into the stable condition *S*, forming a continuation of what is now a 'main stalk'. Further developments are given at the end of the paper.

The successive shifts of the planes once started, may stop suddenly, continuing again after a few seconds, or even stop altogether. As a result of this, one obtains a zigzag crystal, stable and unstable joints alternating, as shown by the photographs in figures 23 and 24, plate 9. They usually pass over to a completely stable, straight crystal, though sometimes 'frozen' as shown by the remarkable example of a large crystal, photographed after it had been taken out of the solution and dried. On the following morning one joint was found to have straightened out during the night (figures 32 to 34, plate 10).

The nature of the 'trigger' which starts the shifting of the lattice planes has not been found. It seemed possible that it might be a thermal 'hot spot' such as the co-operative action that drives the particles in Brownian motion. But what stops the shifting, once started, seems to be even more of a mystery, possibly a blocking of a lattice plane by a minute particle of an impurity. Cooling the solution to zero has not been tried.

A more extensive study of the phenomena was commenced in 1939, and it became at once obvious that it was almost impossible to obtain any results of value if the crystallization was carried on in cells open to the air (Lehmann's method), in which the concentration could not be kept constant, owing to evaporation, and in which the temperature and rate of cooling could not be accurately controlled. I finally succeeded in making small flat micro-flasks of glass, of the form shown in the upper left corner of figure 40 (front and side view), with sides sufficiently flat for observation of the contents with a binocular microscope of low power, and separated by a space of less than 1 mm.

A solution of the acid, saturated at a known temperature (say 60° C), was introduced through the open neck by means of a pipette, with a short delivery tube not much larger than a horse-hair, taking care to keep the fluid away from the neck, which was immediately sealed by a very small flame. The flask could then be heated to any desired temperature in a water-bath, kept at this temperature for a given length of time, to dissolve possible 'seed crystals' left over from earlier trials, and then cooled at a definite rate. It was hoped that in this way it might be possible to obtain only crystals of a single type, instead of a collection of rhombs,

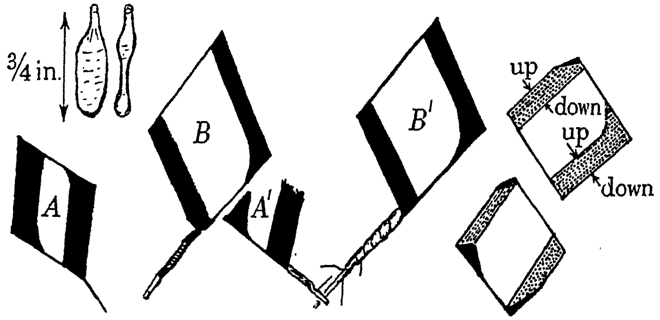


FIGURE 40

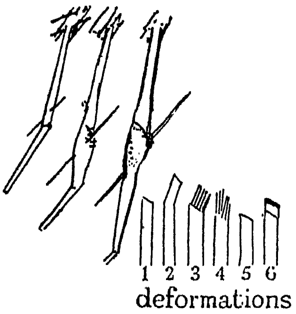


FIGURE 41

TABLE 1

temperature of heating (° C)	time of heating (min.)	yield on cooling
66	10	claws, needles and some rhombs
66	5	claws only, finally a few rhombs
66	30	no crystals after lapse of 1 hr.
	next morning large blocks	
62	10	blocks nearly dissolved; claws form in 5 min.
64	2	rhombs first, then claws and needles
60	4	shook flask and then heated to 61° C
61	6	crystallization started immediately, one claw and needles
65	5	crystallization started at interface between air bubble and solution, needles penetrating latter; after 1 hr. many needles and one claw
	next morning blocks as previously	
66	10	same as last experiment at 65° C
66	1	claws only, formed in 10 min.
66	2	claws in 2 min.
66	2	a few enormous rhombs and some small claws
66	4	rhomb dissolved and after 1 hr. needles from air bubble and one claw
66	not recorded	needles and three claws

prismatic rods, claw crystals and needles radiating from a single point. Plate 9, lower left, not numbered.

The results, shown in table 1, however, were very discouraging, it being almost impossible to duplicate, except by accident, any condition previously obtained. The above series for a single flask is recorded in full, chosen from scores of other series, as it ended with a most astonishing and unexpected phenomenon, which had not been observed in any previous experiment. It will be observed that in each experiment the yield of crystals and their nature depends on the previous treatment of the flask, as well as the time during which it is held at a definite temperature above that of saturation. In some stages the crystallization starts in a few minutes, in others only after a lapse of an hour or two. The observations were made with a rather long narrow flask filled with a solution saturated at 60° C, with a large air bubble at one end, and were carried out over 3 days. When first observed on the second morning, the crystals had united, forming large irregular blocks made up in part from ragged fragments of rhombs.

SPLITTING OF CRYSTALS AND SPONTANEOUS WHIP-LIKE MOVEMENTS OF CRYSTAL FILAMENTS

The above series terminated with a most startling observation. After standing for over an hour a very extraordinary phenomenon was observed, shown in figure 42 *Q* (1, 2 and 3). A rapier-shaped crystal made up of two sections divided by a dark line down its centre opened out slowly from the top as shown by 2, until complete separation occurred at the point, when the two halves instantly whipped around into circular arcs, as shown in 3, the movement occupying less than half a second. A photograph hurriedly made of this event is reproduced in figure 31, plate 9.

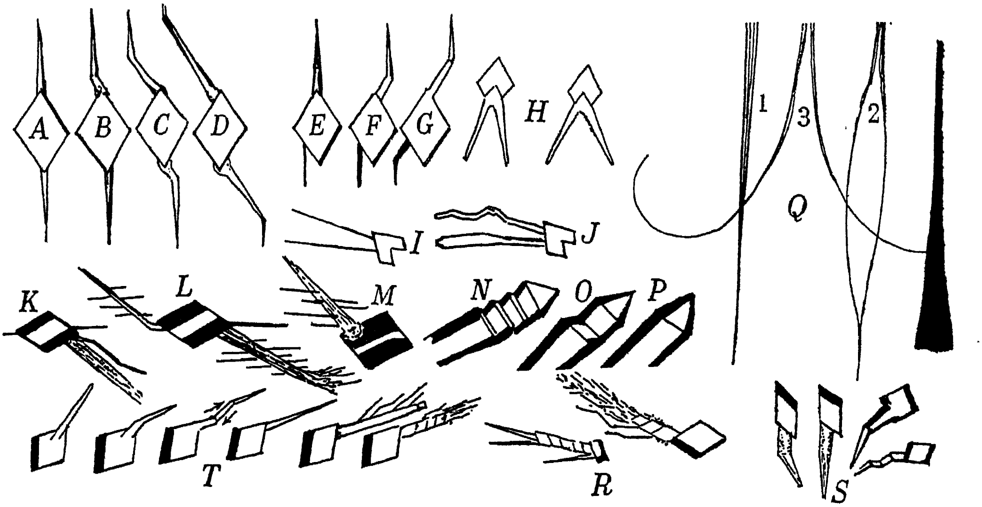


FIGURE 42

Initially the two halves of the rapier crystal must have had a molecular structure not unlike that of two circular arcs of a highly elastic material, bent straight and held back to back by some molecular locking together of the two opposed surfaces. In addition, in order to account for the preliminary 'opening up' of the region between the two halves, which spreads slowly down to the tip, the amount of material under tension, which is involved, must decrease gradually from the top down. A model illustrating the phenomenon can be made from a 10 cm. strip of 16 mm. cinematograph film, cutting it in halves from corner to corner, forming two tapering V's, as shown at right of diagram *Q*, figure 42. These must be given permanent curvatures of small radius (say 2 cm.) by bending them around a fairly hot glass or metal cylinder. Old negative film that has been coiled in a small roll for years usually has sufficient curvature without heating and gives better results. The strips are to be cemented together at their tips with a drop of collodion or 'Duco', when placed in contact back to back, in such a way that the curvatures are properly oriented. If now they are held together between thumb and finger at the thick end, there will be a clear space between them, and on cutting off the tip the strips will promptly whip around. The performance can be shown to a large audience by holding the strips in front of the condenser of a projection lantern and snipping off the cemented tips with a pair of scissors.

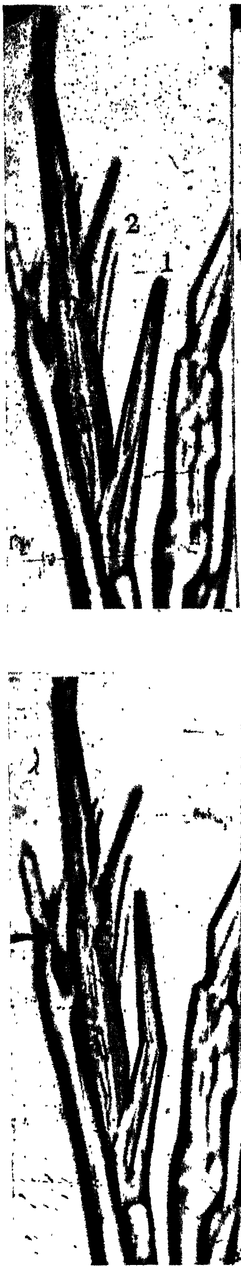
Another crystal, possibly of the same type as the one just described, splits in an irregular manner, breaking into fragments as the process goes on. Near the base, where it is thickest, it has a distinct herring-bone structure (figure 28, plate 9). Two successive photographs of another crystal of this type show the character of the fragmentation. The breaks appear at various points, their extension proceeding in both directions (figures 29 and 30, plate 9).

GENESIS AND LIFE HISTORY OF THE RODS AND CLAWS

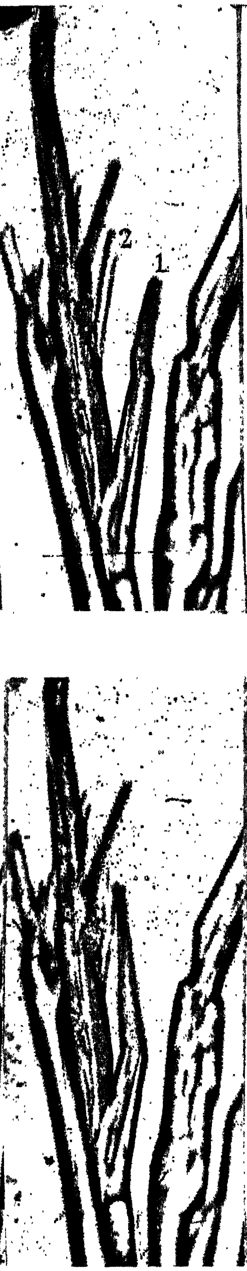
It was only during the past year that an attempt was made to determine the genesis and life history of the rather complicated crystalline forms recorded by the motion pictures. From recent observations made visually, a few of which were photographed and hundreds recorded by quick pencil sketches, it has been possible to form an opinion of how the rods and 'claw'-like groups have developed.

Instead of the micro-flasks a new type of 'cover', superior to the watch-glass method of Lehmann in that it offered a much smaller surface of the solution to the air, was prepared by fusing the four corners of a square 'cover-glass' by touching them with a very minute flame, the small beads formed supporting the glass 0·5 mm. above the surface of the microscope slide. The warm saturated solution was drawn in by capillarity from a drop delivered at the edge by a pipette, and the formation of crystals started usually in a few minutes.

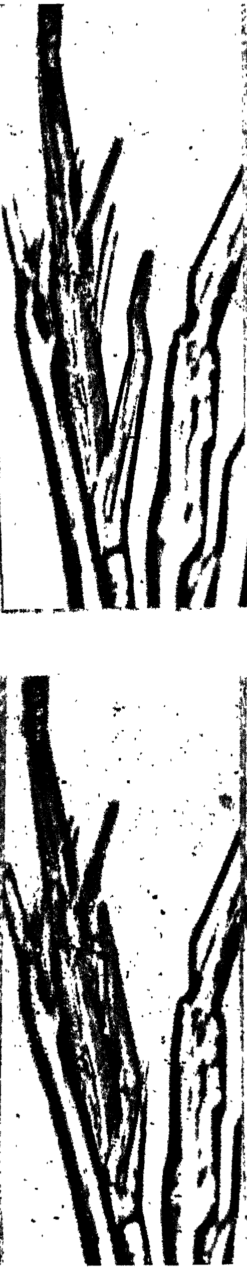
Needles of two types were observed, very narrow ones, radiating from a fixed point, which were inert, as shown at lower left, plate 9, and thicker isolated ones (figure 35, plate 10), which at first sight appeared to be cylindrical spindles, tapering to a point at each end. The spindle immediately develops at its centre a narrow kink of the stable type, shown by the photograph in figure 35, other kinks following in figure 36. Further developments are shown in figures 37 and 38 (a different crystal), giving strong evidence that the crystal is rhombic in cross-section; in



7

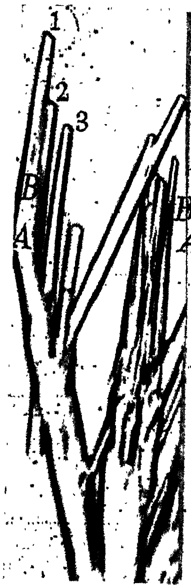


8

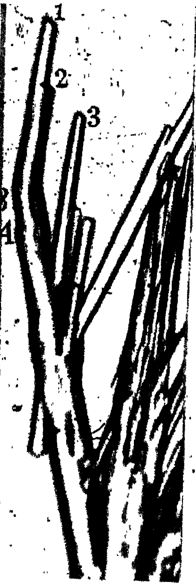


9

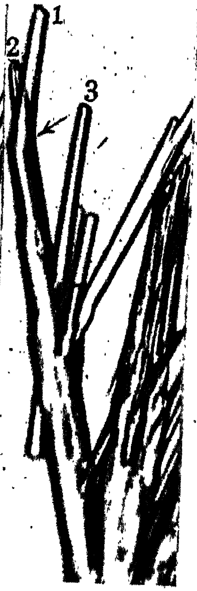
10



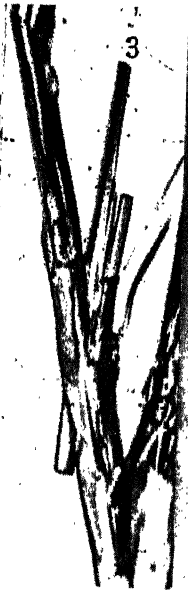
11



12



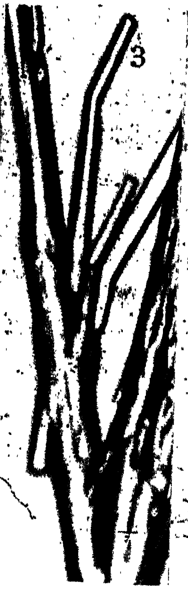
13



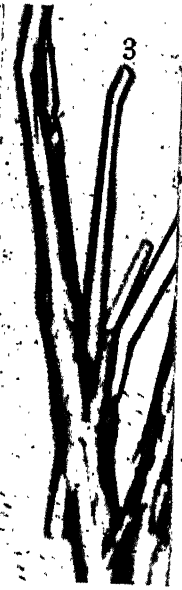
14

15

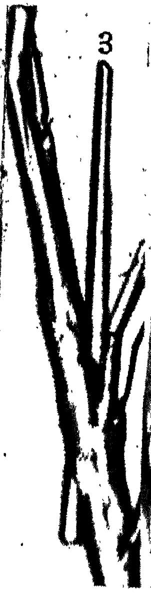
16



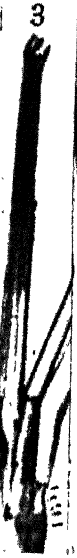
17



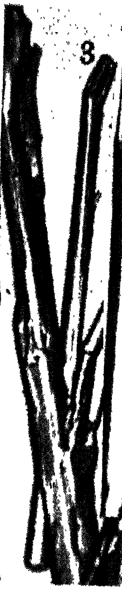
18



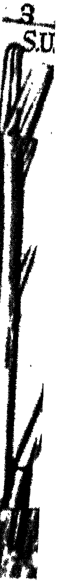
19



20

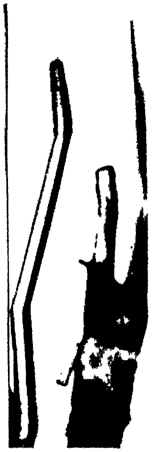


21

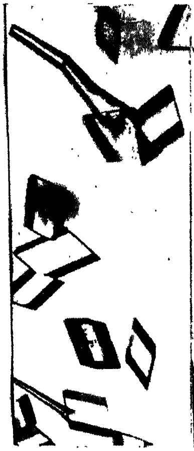


22

23

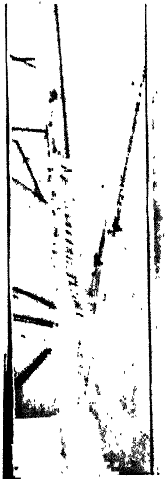

24

25


26

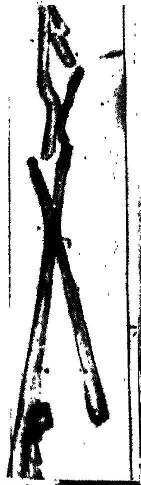
27

28

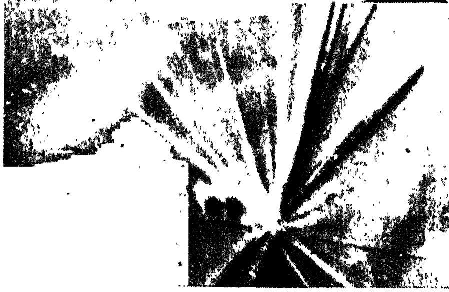


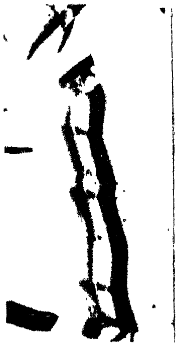
29

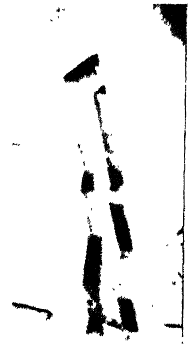
30

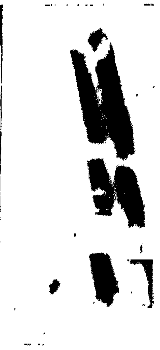


31

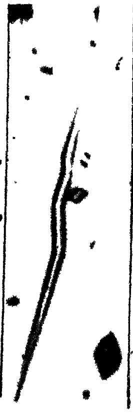


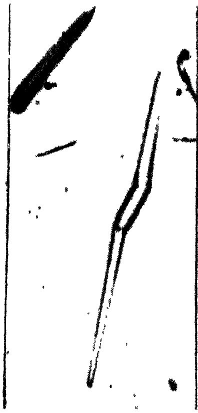

32

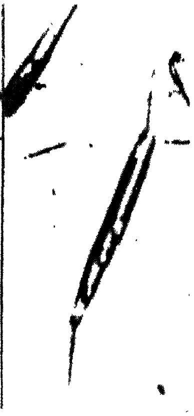

33

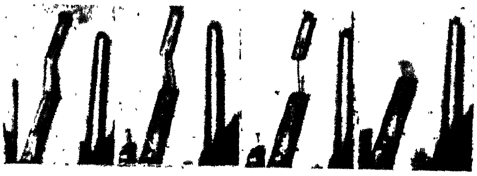

34

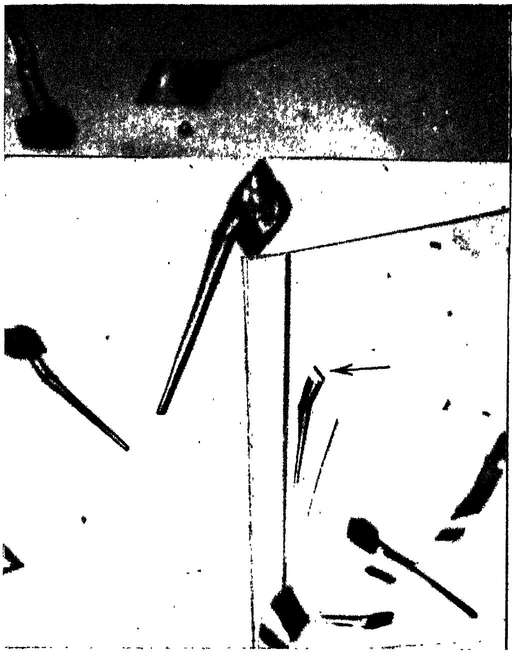
35


36

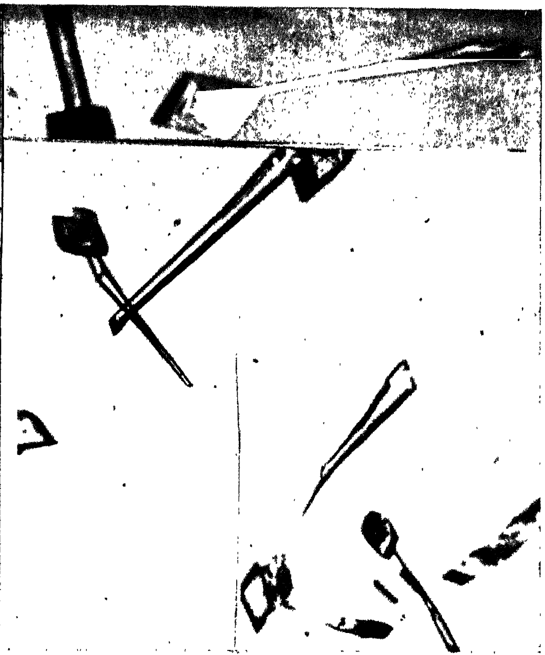

37


38

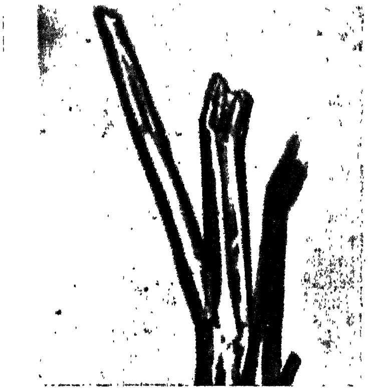

39



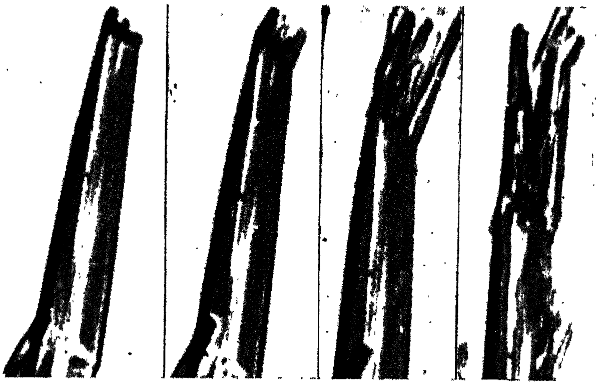
43



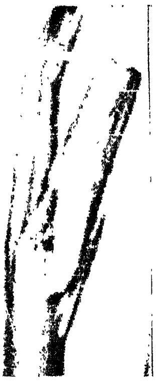
44



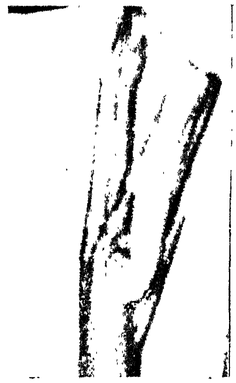
45



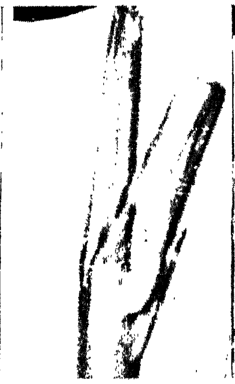
46



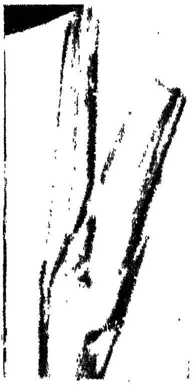
47



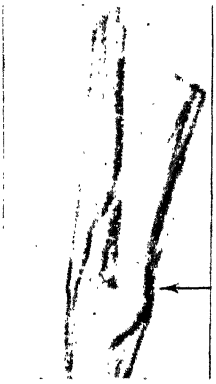
48



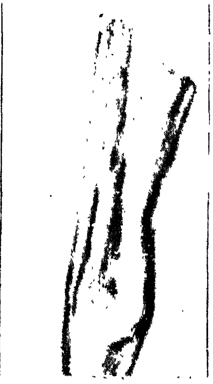
49



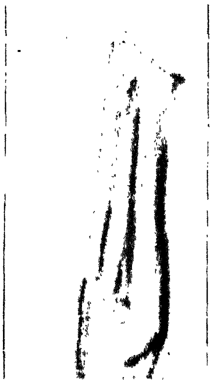
50



51

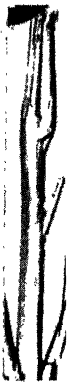


52



53

54



55

fact, a chain of rhombs is actually in the process of formation in figure 38, and many cases have been observed which gives still stronger evidence that what I have called the main stalk has an internal structure consisting of a chain of rhombs arranged end to end, as shown by figure 42 *R*.

INTERNAL STRUCTURE OF STALK AND ITS DIRECTIVE ACTION ON THE GROWTH OF RODS AND CLAWS

We can regard the thickened stable part of the spindle with its internal structure of rhombs, just mentioned, as the early stage of a main stalk, and it appears certain that it is this structure that causes the unstable rods to grow out, always at the same angle of obliquity and along a straight line on one side of the stalk. Needles and rods have been observed sprouting along another straight line on the opposite side, at the same angle of obliquity, but pointing in the opposite direction, as shown in figure 42 *L*. Figure 27, plate 9, found on the cinema film, shows the oppositely directed rods on the main stalk. One case was observed in which a well-isolated stable rod thickened at one point and sprouted a pair of needles one on each side at the same obliquity and parallel to the unstable joint at the free end of the rod (figure 41). The spindles start as extremely narrow threads which thicken more rapidly at the centre, producing the double tapering referred to previously.

UNSTABLE RODS BASED ON RHOMBS

Under certain conditions of saturation and cooling, some of the material crystallizes in the form of thin diamond-shaped plates, with an unstable spindle growing out from the vertex of one or both of the acute angles of the diamond, and parallel to its longer diagonal. Two sequences are shown in figure 42, *A* to *D*, in which a half-spindle grows from each acute vertex, and *E* to *G*, in which the growth is from a point on one of the sides. In both cases deformation starts at the point of attachment, and runs along the spindle, its base appearing to crawl around the apex until entirely on one side in *D*. Whether this puzzling performance results from molecular shifting, or a combined process of solution and recrystallization, is uncertain, and requires investigation with polarized light. If the upper spindle turns to the left or counter-clockwise, the lower one always turns to the right, also counter-clockwise. In sequence *E* to *G* the lower unstable growth is based on one of the sides, but parallel to the one at the apex, and the direction of rotation is clockwise for both.

The next sequence *H* of two figures shows what might be termed twin rods growing from the lower apex of a rhomb, which opened suddenly like a pair of compasses, the branches moving so rapidly that the movement resembled that of two rigid bodies. Motion pictures would be necessary for recording the small details of the movement. On numerous occasions one observes thin crystals, similar to those shown at figure 42, *I* and *J*, which may be twins, figure 42 *I*, carrying two long unstable needles, which, at the end of 13 min., had developed as shown by *J*. The increase in the diameter of a rod with increasing distance from the point of support is seen very frequently, and photographs of such crystals are shown in the group in figure 44, plate 11, taken some minutes after figure 43. The left-hand picture of the upper pair showed no thickening of the very thin rod based on the apex of the diamond-shaped under-surface of the well-developed rhomb or possibly on its dark side, while the gradual thickening and deformation into the stable condition appear in the right-hand picture though with a rather ragged terminus. On the other hand, in the lower group of figure 43, a similarly developed deformation is crowned with a beautiful rhomb (marked by arrow). In figure 44, taken a few minutes later, no increase in the distance between the terminal rhomb and the one to which it is attached has taken place, though the joint separating the stable from the unstable part has disappeared. Similar conditions must have occurred frequently, as many crystals, such as those shown in figure 42 *S*, are recorded in my notes, though their genesis was not determined, and it was assumed that the prism was an outgrowth from the rhomb instead of vice versa. Several cases, however, were found on the moving-picture film about which no doubt could exist as to which element formed first. In figure 42, *K* and *L* are copies of the same rhomb at different stages of its growth; which record a minute unstable needle based on the left-hand apex and parallel to the long axis of the diamond *K*, which presently on *L* has stabilized itself parallel to the side, and has sprouted needles, while its base is still unstable and parallel to the long diagonal of the diamond. Unfortunately, the intervening steps did not appear on the film owing to a shift of the field of view. It seems to be the reverse of what happened at the bottom of diamond *C*. Much more advanced growths appear on the other end of the rhomb.

Crystal *M* showed clearly that the stable, conical rod was based on the upper surface of the rhomb, and the sequence *T*, below *KLM*, also drawn from the film, shows where an unstable joint has been trapped between two stable ones, the dark cross-lines approaching each other, as shown by the arrows. The last one of this group shows that the small rhombs which have formed in the stalk are lined up with the larger one.

The most convincing case of all, however, is shown in figure 40, which figures two extremely large rhombs *A* and *B* in close proximity, on a film made by Dr R. P. Loveland of the Eastman Co. with material which I had sent to Dr Mees. These were copied, as they showed the curvature at certain corners better than any of the photographs. Interesting also were the growths of a needle and a jointed rod which exhibited a row of very minute white spots. These in the second picture, *A′* and *B′*, had acquired the rhombic outlines mentioned in previous cases.

Adjacent drawings of two rhombs which were made from visual observations show the small oblique faces which frequently form on the corners of the crystal.

DEFORMATION OF RODS AND 'DACTYLS'

When the growth of an unstable rod has halted momentarily in the stable state with a sloping summit as shown in figure 19, plate 8, one of three things may happen: it may develop further as an oblique growth, also with a summit which is not perpendicular to the sides, as shown in 1 and 2 in the group in figure 41 marked 'deformations', or, if thicker, and with a larger surface at the top, shoot out a number of unstable finger-like rods or 'dactyls', perpendicular to the sloping summit, which eventually set themselves parallel to the lower part of the rod, as shown by 'deformations' 3 and 4 in figure 41 and also in figures 20 and 21, plate 8. Under some circumstances, however, further growth proceeds only by a thickening of the stable rod and the formation of a rhomb, one edge of which is seated on the slanting summit shown by 'deformations' 5 and 6, figure 41. A sequence of four frames from the motion pictures showing the growth of dactyls made with the 32 mm. film from the Berlin material, is reproduced in figure 46, plate 11.

RELATIVE SOLUBILITY AND GROWTH OF STABLE AND UNSTABLE SEGMENTS OF THE CRYSTALS

After the growth of crystals of mixed type has ceased, the fine needles dissolve first, then the unstable rods and the unstable segments of jointed rods. The stable segments increase in thickness during this process by utilizing the material going into solution from the more soluble elements. After 8 or more hours all of the soluble material has been taken up by the rhombs, everything else having disappeared, and finally everything may be captured by two or three of the largest rhombs.

Four photographs of a rod having three joints, the middle one unstable, are shown in figure 39, plate 10. The first was taken after the growth of the crystal was complete, the next three at appropriate intervals showing the gradual solution of the middle section and the continued growth of the stable ones, the upper one finally disappearing when the fine thread supporting it gave way and it dropped out of focus. Figure 38, plate 10, also indicates that the lower unstable end of the spindle is passing into solution.

LATER STUDY OF THE DEFORMATIONS

The further developments of crystal 3 in figure 19 of plate 8, after reaching the stage shown in figure 22 of the same plate, are shown in plate 12, where the unstable oblique crystal is larger than the vertical stable one on which it is based. Figures 47 to 51, plate 12, show only the continuous increase in the length of the stable crystal by the formation and deformation of dactyls, which in figure 47 give the impression of forming a hollow tube, the upper portion of which has deformed to the stable type in figure 48, while in figure 49 and the following figures the entire crystal has become stable. These figures are highly enlarged from the 32 mm. motion picture film.

No deformation of the oblique unstable crystal appears until figure 51, where the base has become parallel to the stable crystal at the point indicated by the arrow. From here on it becomes a little difficult to decide on just what happens. The deformation into the stable configuration moves up the crystal in figures 52 to 54, which show clearly the gradual passage of the crystal from the oblique to the vertical position, and its final position *behind* and parallel to the stable crystal.

The transition from 51 to 52 is the one which is difficult of interpretation, since the continuity of the edge of the stable crystal appears to be interrupted (optically), though becoming continuous again in figure 53. Note, however, that this same interruption appears in figures 47 to 52. Figure 55 was the final stage recorded on the film, though less highly magnified.

REFERENCES

Frick, J. 1890 *Physikalisch Technik*, 6th ed. (ed. O. Lehmann). Braunschweig: Friedrick Vieweg.

Frick, J. 1905 *Physikalisch Technik*, 7th ed. (ed. O. Lehmann). Braunschweig: Friedrick Vieweg.

The oxidation of copper and the reactions of hydrogen and carbon monoxide with copper oxide

BY W. E. GARNER, F.R.S., T. J. GRAY AND F. S. STONE

Department of Inorganic and Physical Chemistry,
The University of Bristol

(*Received* 6 *September* 1948—*Read* 17 *March* 1949)

[Plate 13]

An investigation has been made of the activation of copper by successive oxidation and reduction with hydrogen. Reduction with carbon monoxide causes deactivation of the surface.

The heats of adsorption of carbon monoxide and oxygen and the heat liberated during the catalytic reaction of carbon monoxide and oxygen on a cuprous oxide film formed on metallic copper have been measured at room temperature. The kinetics of the reactions have been studied.

The electrical conductivity of thin films of oxides has been measured during the reduction with hydrogen, and a study made of the process of embrittlement, whereby the surface is activated. The effect of the adsorption of gases on the conductivity of the Cu_2O-CuO surface has been investigated. Oxygen enhances the conductivity and carbon monoxide and hydrogen depress it. The reaction between carbon monoxide and oxygen has been followed at room temperature by conductivity measurements and the results have been employed to interpret the nature of the adsorption, the kinetics of the reaction, and the deactivation of the surface by carbon monoxide.

INTRODUCTION

The mechanism of the oxidation of copper over the high-temperature range, 800 to 1000° C, has been explained by Wagner and his co-workers in quantitative terms (Wagner 1933; Wagner & Grünewald 1938). They postulate that the oxide layer, which is composed of Cu_2O in this temperature range, is a defect structure containing vacant lattice sites normally filled by Cu^+ ions (Schottky defects). The oxide thus possesses oxygen ions in excess, and the excess of oxygen in the lattice over the stoichiometric relationship is dependent on the pressure of oxygen.

During oxidation, the oxide film grows by the diffusion of Cu^+ ions, together with an equivalent number of electrons, across the film from the metal to the surface of the oxide, where they combine with oxygen.

Although the theory gives a quantitative explanation of the phenomena occurring at temperatures above 500° C, there are considerable difficulties in the interpretation of the results for the oxidation of copper at lower temperatures, 0 to 300° C, where the parabolic law is not, in general, obeyed (cf. Lustman 1942). The rate of oxidation over the lower temperature range is a structure-sensitive property, depending on the treatment of the metal before oxidation.

The reduction of the oxide with hydrogen or carbon monoxide might be expected to be the reverse of the oxidation process. If this occurred, the removal of oxygen ions from the external surface of the film as water or carbon dioxide would set free electrons which, with their associated copper ions, could then diffuse across the oxide layer to build up the metallic phase. If, however, the chemical processes on the surface are very rapid, the surface concentration of electrons may become so great that nuclei of metallic copper will be formed near the oxide surface. This will produce a copper-copper oxide interface much closer to the site of the surface reaction than the original interface, and this would grow preferentially. It is shown in this paper that such nucleation does, in fact, occur during the reduction of a thin film of cuprous oxide on copper (see p. 297).

It has long been known that the activity of a copper surface can be enhanced by repeated oxidation and reduction with hydrogen. The surface of the metal is roughened in the process, and the increased activity of the surface is due to the roughening that occurs. Ransley (1939) and Rhines & Anderson (1941) have shown that the lattice of the metal is disrupted by the reaction of hydrogen with inclusions of oxide embedded in the copper metal. The water formed builds up to a high pressure, and in its escape along grain boundaries and cleavage planes the metal lattice is disrupted. This is the well-known embrittlement of copper. The reduction of oxide films with hydrogen might therefore be expected to give metal in a porous form.

The surface of copper and copper oxide can be so activated that it can act as a catalyst at room temperature for the reaction between carbon monoxide and oxygen (Jones & Taylor 1923). Activated copper oxide either alone or in admixture with other oxides has been used in gas masks for the removal of carbon monoxide from the air (Lamb, Bray & Frazer 1920). It is of interest to study the reactions on activated copper and its oxides, for occurring as they do at room temperature, it is possible to employ in the investigation methods which are not readily applicable at higher temperatures. In the reaction between carbon monoxide and oxygen, these gases will be adsorbed and oxygen ions, and probably carbonate ions, formed as intermediate states. Carbonate ions have been demonstrated as intermediates in the case of the reduction of oxides of the transition elements (Garner 1947) and may be produced in this case also.

For the investigation of such systems, in which the intermediate states may radically change the characteristics of the surface film of the catalyst, measurement of electrical conductivity should prove valuable, especially as the oxide is itself a semi-conductor. Heat of adsorption measurements are also useful, since from such measurements it would be possible to make deductions as to the structure of the intermediate states. In this investigation, therefore, two separate reacting systems have been studied under such conditions that either electrical conductivity or heat of adsorption measurements can be made simultaneously with changes in the pressure of the gases.

The research is divisible into three parts, the first dealing with the activation of copper oxide, the second with measurements of heats of adsorption and reaction on the activated surface, and the third with measurements of changes in superficial conductivity during the adsorption and reaction of the gases. The data have been employed to explain the nature of the reaction between carbon monoxide and oxygen on copper oxide.

THE ACTIVATION OF COPPER OXIDE BY OXIDATION AND REDUCTION

Reduction by hydrogen

The earliest work on the reduction of copper oxide was carried out by Wright, Luff & Rennie (1878), who observed an induction period which was the longer the lower the temperature. The existence of an induction period has been confirmed by several workers, notably Pease & Taylor (1921, 1922), who attributed the induction period to the initial absence, the formation and the growth of metallic nuclei. Also, Palmer (1923), Hinshelwood (1923), Juliard (1932) and Hasegawa (1943) have come to similar conclusions. The reaction during the reduction thus occurs preferentially at the boundary of the metal and its oxide.

Reduction by carbon monoxide

References in the literature to the reduction of cupric oxide by carbon monoxide are not very numerous. Jones & Taylor (1923) found that carbon monoxide produced nuclei of metallic copper on cupric oxide, as was the case with hydrogen, and that oxygen present in the carbon monoxide inhibited the reduction by oxidizing the copper nuclei as fast as they were formed. If, however, oxygen was admitted to the reducing gas after the nuclei had been formed and well established, then the carbon monoxide reacted with the oxygen by contact catalysis.

Experimental

Oxidation and reduction

A strip of pure copper 2 by 1 cm. and 0·0125 cm. thick was degreased with carbon tetrachloride and polished by rubbing with finely ground 'Analar' precipitated cupric oxide. It was then annealed at 600° C in an atmosphere of hydrogen. It was subjected to successive oxidation and reduction at 300° C, 0·08 ml. of oxygen being employed at an initial pressure of 0·08 cm. The reactions were carried out in a cylindrical Pyrex vessel, which possessed a plane window for microscopical and visual observation. A Zeiss film camera was employed for the microphotographs. The vessel was wound with nichrome wire and heated electrically. The temperature of the interior of the vessel was measured by means of a thermocouple. Purified dried gases were used.

The activity of the metal was found to increase with successive oxidation and reduction, and the maximum activity was reached after four or five repetitions of the processes. The surfaces which were initially smooth became much roughened owing to the formation of blisters (figure 1*a*, plate 13). This roughening is believed to be due to the explosive removal of water vapour from oxide included in the metal (Ransley 1939). The oxidation did not obey a parabolic law in the initial stages of the reaction, although curves for the fall in pressure of the oxygen against $\sqrt{t}$ showed this law tends to be obeyed in the later stages of the reaction. The curves show that the rate of oxidation increases during the first four or five experiments. The processes occurring are evidently very complex, which is understandable in view of the roughening which occurs during the activation process. It was found that at 300° C the rate of oxidation was not appreciably affected by increasing the pressure four times, and it was unaffected by interruption of the oxidation process by withdrawing and reintroducing the gas. There was only a slight increase in the rate of oxidation on increasing the temperature by 50° C.

Although no interference colours were obtained in the first oxidation, in all the subsequent oxidations bright colours were observed. It was possible to employ these to give some measure of the roughness of the surface when the copper was activated. The thickness of the oxide film was calculated from the observed colour by the method of Constable (1927, 1928). The values are compared in table 1 with those calculated from the fall in pressure, assuming that the area of the specimen was the external area and that the density of the oxide layer was that of Cu_2O. The values are the means of several oxidations. The results in table 1 give a roughness factor of approximately 10.

TABLE 1

	thickness of film (Å)	
colour	from interference colours	from pressure
purple	460	4120
blue	530	5870
green	880	9890
yellow	1000	16500

Hydrogen readily reduces the oxide film at 300° C, but there was no evidence of the occurrence of an induction period until the fourth reduction, after which a marked induction period was found. This is in accord with the work of earlier investigators. The plot of log Δp against log t during the induction period (figure 2) gives straight lines of which the slope is 2·0, and this supports the view that a limited number of nuclei are formed immediately and that these increase in size at a linear rate (cf. Wischin 1939). The surface lightens in colour during the reaction, but the individual nuclei were too small to be seen with the magnification used.

Towards the end of the reduction, visible nuclei of a different type appear (figure 1*b*, plate 13), which also grow at a linear rate with time. These are believed to be due to the sintering of the small copper crystals formed early in the reaction to give large crystals of metallic copper.

In some experiments the pressure of the hydrogen was changed, during the reduction, and the rate of reduction was found to be roughly proportional to the hydrogen pressure. The temperature was changed from 300 to 350° C, and from the temperature coefficient of 2·25 for 50° C an approximate value for the activation energy of 12 kcal. was calculated.

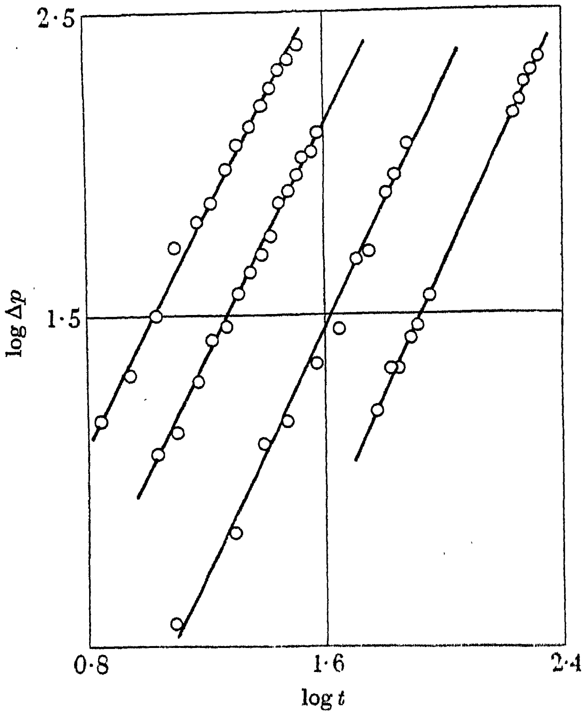


FIGURE 2. Reduction by hydrogen.

Reduction with carbon monoxide

The reduction did not give an observable induction period, yet metallic copper appears very quickly on the surface of the oxide. The nuclei are so numerous that they soon coalesce to give a coherent film of metallic copper and a sandwich of cuprous oxide is thereby produced between two layers of copper metal.

The rate of reduction by carbon monoxide of an oxide film which has been previously activated with hydrogen is very similar to that obtained with hydrogen. As a result of the reduction, however, the copper loses much of the activity which has been conferred on it by the embrittlement process. It is not possible to carry out two successive reductions with carbon monoxide at 300° C, for the second reduction soon becomes very slow and erratic, and finally ceases. It is necessary to reactivate the surface by reduction with hydrogen before the reduction with carbon monoxide can again be effected. The rate of reduction by carbon monoxide, similar to that by hydrogen, is approximately proportional to the pressure of the gas.

The destruction of the activity of the copper by carbon monoxide, which has been observed even at lower temperatures, will be discussed later in the paper.

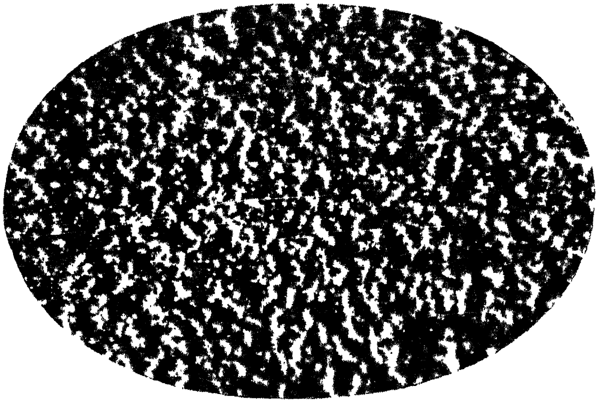


FIGURE 1 (*a*). Formation of blisters on the surface of copper oxide during reduction by hydrogen (magn. × 75)

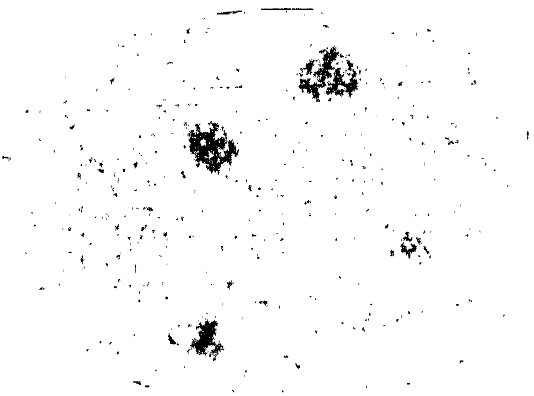


FIGURE 1 (*b*). Nuclei of copper on a film of copper oxide during later stages of reduction by hydrogen (magn. × 75)

(*Facing p.* 298)

HEATS OF ADSORPTION ON COPPER OXIDE

Calorimeter

Measurements were made in a thermocouple adsorption calorimeter similar in principle to those employed in earlier work on adsorption on oxides (Garner & Veal 1935). The adsorbent consisted of pure metallic copper sheet, $1\cdot25 \times 10^{-3}$ cm. thick, cut in the form of disks and threaded on a thin glass tube (*D*, figure 3). 1000 disks, 2 cm. in external diameter and of total area approximately 5,700 cm.^{2}, weighing 27·5 g., were used. They were not touched by hand during cutting or assembly. The disks were compressed until they occupied a height of 8 cm., and were held at either end by glass crosses *E* and *F*. A platinum-platinum rhodium thermocouple passed down the tube *D* to a depth of 4·5 cm., and the junction was embedded in the copper. The method adopted in bringing the thermocouple wires out of the calorimeter will be seen from the diagram. There are holes for the inlet of gas in tube *D* not shown in the diagram. The copper disks were enclosed in a thin cylindrical glass bulb which was connected by the tube *D* to the burette holding the gas, of which the heat of adsorption was to be measured. The total weight of the bulb and of the other Pyrex glass used in the calorimeter was 4·07 g. The water equivalent of the calorimeter, calculated from the specific heats of the components, was 3·37 g. This was subsequently checked against the heat of combustion of carbon monoxide occurring catalytically on copper oxide and good agreement found.

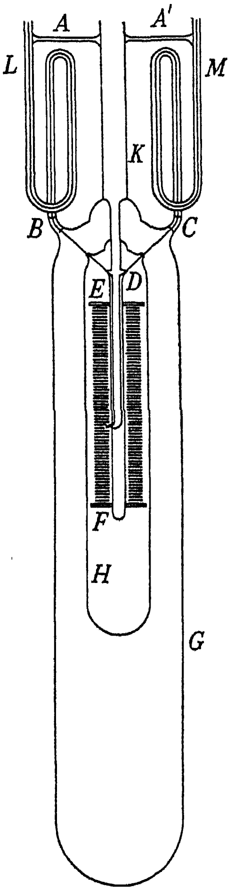


FIGURE 3. Calorimeter.

The glass bulb was enclosed in a larger glass cylinder, and the space between could be evacuated simultaneously with the inner bulb. To diminish radiation losses the external surface of the inner bulb and the internal surface of the outer bulb were platinized.

During the measurement of heats of adsorption the calorimeter was immersed in water in a Dewar vessel, and the cold junction of the thermocouple was placed in a similar Dewar. Beckmann thermometers were used for the purpose of calibration. The calorimeter and the cold junction were placed in a segmented heat-insulated box. A mercury switch (Garner & Veal 1935) was employed to connect the thermocouple to a Kipp and Zonen Z.C. galvanometer, the deflexion of which was recorded photographically. The arrangements were very similar to those employed in previous investigations and need not be described in detail. The sensitiveness of the thermocouple system was 14 cm./° C.

The main loss of heat in this type of calorimeter occurs by radiation and convection, and loss by conduction along the inlet tube and the thermocouple wires is negligible. The Newton coefficient k from $-dT/dt = k(T-S)$, where T is the temperature of the thermocouple junction and S the temperature of the outer walls of the calorimeter, was determined from cooling curves for varying pressures in the inner bulb, a hard vacuum being maintained in the outer bulb. k(deg.min.$^{-1}$/° C) varied from $-0\cdot025$ at $p = 10^{-5}$ cm. to $-0\cdot039$ at $p = 10^{-1}$ cm., which values are somewhat greater than the value of 0·020 obtained in a similar calorimeter (Garner & Veal 1935).

The vacuum system comprised two gas burettes, two expansion bulbs, two McLeod gauges, a Toepler pump, a mercury-vapour trap, connected to a Hyvac pump and mercury-diffusion pump. The tap connecting the vacuum system to the calorimeter was never opened unless there was liquid oxygen around the mercury-vapour trap, so that the introduction of mercury vapour to the copper surface was avoided. Apiezon grease was used in the tap. The small vapour pressure of this grease did not affect the phenomena observed to an appreciable extent, since these were confirmed in a greaseless vacuum system used to measure the electrical conductivity. The McLeod gauges enabled pressures to be read from 4 mm. to 10^{-5} cm. The volume of various parts of the apparatus was determined, and the theoretical pressure drop on introducing gases from the McLeod system to the calorimeter was calculated.

During an experiment curves for the fall of pressure with time and for the rate of liberation of heat were obtained; and by nomographic methods the integral and differential heats of adsorption were evaluated. Measurements were made of the carbon dioxide liberated during the reaction and on baking out by condensation in a liquid oxygen trap.

Activation of the copper

A thin layer of oxide was formed on the surface of the disks (c. 500 Å) by oxidation at 150° C. After the first oxidation, the surface adsorbed at room temperature only 0·015 ml. (N.T.P.) carbon monoxide at a gas pressure of 0·03 cm. The film was reduced by hydrogen and then re-oxidized. The carbon monoxide adsorption was increased, but not very materially. Repeated oxidation and reduction of the surface, however, caused the adsorption of carbon monoxide to increase to 0·060 ml. There are thus $1\cdot6 \times 10^{18}$ molecules of cross-sectional area $1\cdot2 \times 10^{-15}$ cm.^{2}, which, if tightly packed, would saturate a surface area of $1\cdot3 \times 10^{3}$ cm.^{2}. The external area of the metal is $5\cdot7 \times 10^{3}$ cm.^{2}, and there is a considerable roughness factor, so that the adsorption is occurring on only a small fraction of the area of the oxide. Usually after oxidation the apparatus was baked out in hard vacuum at 150° C for 1 hr. before cooling to room temperature for the adsorption measurements. It was not baked out for a longer period, in order that as much as possible of the adsorbed oxygen should be retained by the surface (see below).

There was a small 'gas effect' with this calorimeter, especially when using oxygen, so that helium was always added to the inner bulb to remove temperature gradients before admitting the adsorbate.

Oxygenated and deoxygenated surfaces

When cuprous oxide has been exposed to oxygen either as a result of oxidation at 150° C, followed by pumping for an hour, or as a result of adsorption of oxygen at room temperature, the surface becomes charged with oxygen in a mobile form, which increases the surface conductivity. The oxygen is decreased during continued pumping at 200° C, but it is firmly held at room temperature. The behaviour of the surface towards adsorbed gases depends on the quantity of mobile oxygen on the surface.

A surface which contains adsorbed mobile oxygen is said to be 'oxygenated', and a surface from which the oxygen has been removed by carbon monoxide to be 'deoxygenated'. Deoxygenation was brought about by treatment with carbon monoxide at room temperature and baking out at 200° C, which removes the oxygen as carbon dioxide. A deoxygenated surface can be oxygenated by treatment with oxygen at room temperature, but it is not so active as freshly oxidized material, since in the treatment with carbon monoxide the surface is partially deactivated. After two treatments with carbon monoxide, the surface loses its power to adsorb mobile oxygen and becomes inactive.

Adsorption of carbon monoxide

Deoxygenated surfaces

0·06 ml. of carbon monoxide is adsorbed on a deoxygenated surface at $p = 0{\cdot}03$ cm., and the heat of adsorption is 28 kcal./mol. The heats in this condition are very reproducible, and values in successive experiments which have been obtained are 28·5, 28·4 and 28·2 kcal./g.mol. The adsorption is practically complete within 5 min. and the heat remains constant throughout the period of the adsorption. There is no evolution of carbon dioxide during the adsorption of carbon monoxide on the deoxygenated surface. After removal of the unadsorbed gas and baking out at 100 to 150° C into a closed evacuated system, about 15 % of the adsorbed gas is desorbed as carbon monoxide, together with some carbon dioxide. If the temperature is raised to 200° C more carbon dioxide is desorbed and the 15 % of carbon monoxide now reacts with the oxide and condenses in the liquid oxygen trap as carbon dioxide. Only about 95 % of the adsorbed gas has been recovered on desorption, possibly because some of it is evolved as carbon monoxide in the preliminary evacuation of the oxide prior to baking out.

Oxygenated surfaces

The heat of adsorption on an oxygenated surface varies between 28 and 47 kcal., according to the quantity of mobile oxygen on the surface. Carbon dioxide is evolved *during* the adsorption of the carbon monoxide, and the exaltation of the heat depends on the amount of carbon dioxide formed. In one experiment which gave a heat of 35 kcal./g.mol., 0·070 ml. of carbon monoxide was removed, 0·026 ml. carbon dioxide was liberated during the run, and a further 0·041 ml. carbon dioxide obtained on baking out up to 200° C. It will be noted that in this experiment, 0·013 ml. of mobile oxygen has been removed to give carbon dioxide at room temperature.

The curves for the change in pressure with time for adsorption on an oxygenated surface are compared with those for a deoxygenated surface in figure. 4. During the first 2 min. the quantities of gas adsorbed are similar in both cases, but whereas in the case of a deoxygenated surface an equilibrium value is established in about 5 min. there is a slow reaction in the case of the oxygenated surface, which continues for 25 to 30 min. and is then completed. During the adsorption on the oxygenated surface carbon dioxide was evolved. An analysis of the heat and pressure curves shows that only a small fraction of the carbon dioxide is evolved during the first 5 min., the bulk of the carbon dioxide being produced after $t = 5$ min. The slow displacement of mobile oxygen by carbon monoxide is also brought out by the conductivity experiments (figure 10).

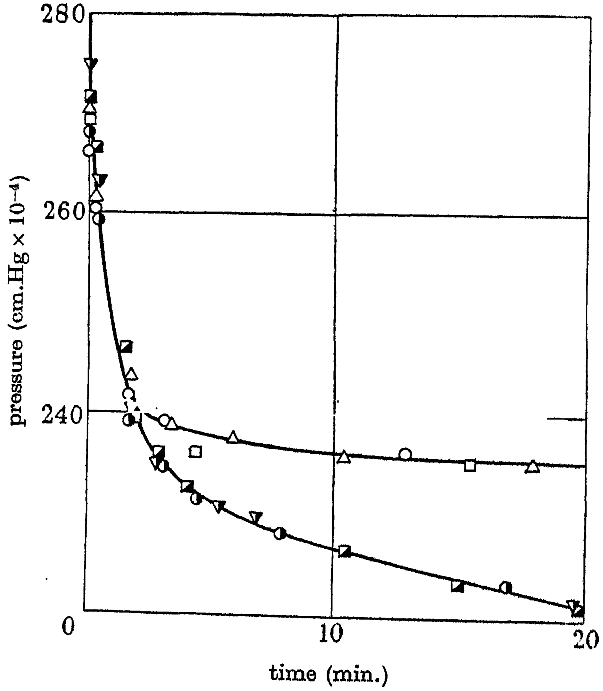


FIGURE 4. Adsorption of carbon monoxide. △, □ and ○ = deoxygenated surfaces, ▼, ◪ and ◑ = oxygenated surfaces.

The main characteristics of adsorption of carbon monoxide on an oxygenated surface are

(1) initial rapid adsorption of carbon monoxide;

(2) a slow reaction between the adsorbed mobile oxygen and adsorbed carbon monoxide to give carbon dioxide, which evaporates;

(3) adsorption of carbon monoxide in the spaces on the surface left vacant in (2).

Reaction of carbon monoxide with oxygen

If a stoichiometric mixture of carbon monoxide and oxygen be admitted to a deoxygenated surface, there is an initial rapid adsorption of carbon monoxide, followed by steady decrease of pressure due to the catalytic combination of the two gases to give carbon dioxide. If the surface was saturated with carbon monoxide prior to the admittance of $CO + \frac{1}{2}O_2$, the rapid initial uptake of gas was absent and only the slow process occurred.

The pressure-time curve for the adsorption is similar to that for carbon monoxide on the oxygenated surface (figure 4) except that the downward trend after $t = 5$ min. is somewhat steeper and continues for as long as the unreacted gas is present. At the end of the run the surface contains adsorbed carbon monoxide which can be pumped off as carbon dioxide on baking out at 200° C.

The initial heat of adsorption at $t = 2$ min. is 20 to 30 kcal., followed by a strictly linear differential heat per mol. of gas $(CO + \frac{1}{2}O_2)$ removed of 45 kcal. (see table 2). The initial heat is mainly due to the adsorption of carbon monoxide, and the heat of 45 kcal. can be quantitatively accounted for by the occurrence of the reaction $2CO + O_2 \rightarrow 2CO_2 + 2 \times 68$ kcal.

TABLE 2

time in min.	0	1	1½	2	3	4	6
integral heat in kcal./mol.	—	20	27	32	—	—	—
differential heat in kcal./mol.	—	32	44	45*	45*	45*	45*

CO adsorbing
←——————→

←$CO + \frac{1}{2}O_2$ reacting→

* Denotes the heat per mol. of gas mixture $CO + \frac{1}{2}O_2$.

The following calculation shows the quantitative nature of the agreement with the above mechanism. In one of the experiments, 0·073 cal. had been liberated at the fifth minute. This was partly due to the adsorption of carbon monoxide ($Q = 28$ kcal.) and partly to the formation of some carbon dioxide ($Q = 67{\cdot}5$ kcal. per mol. CO). Since 0·029 ml. of carbon monoxide was found to be adsorbed at the end of the reaction, one may assume that this amount of gas is on the surface at $t = 5$ min. This will have liberated $0{\cdot}029/22{\cdot}4 \times 28 = 0{\cdot}036$ cal. The observed volume decrease is, however, 0·048 ml., so 0·019 ml. is removed by the $(CO + \frac{1}{2}O_2)$ reaction. Thus, the heat liberated in the formation of carbon dioxide is $0{\cdot}019 \times \frac{2}{3} \times 67{\cdot}5/22{\cdot}4 = 0{\cdot}039$ cal. The total heat liberated should, therefore, be $0{\cdot}036 + 0{\cdot}039 = 0{\cdot}075$ cal., as compared with the experimental value 0·073 cal. Similar calculations with different runs and different values of t also give agreement within 5 %.

If the hypothesis that the reaction occurs on a carbon monoxide-saturated surface be correct, then the rate of reaction will be proportional to the partial pressure of the oxygen present, $-dp_{O_2}/dt = \frac{1}{3}kp_{O_2}$, where p_{O_2} is the partial pressure of oxygen. The integrated form of the equation,

$$\log_{10} p_{O_2} = -\frac{kt}{2{\cdot}303 \times 3} + \text{const.},$$

can be employed directly for cases where the surface is saturated with carbon monoxide, before admitting the reaction mixture. A straight line is obtained, for the plot of $\log_{10} p_{O_2}$ against time, showing that the equation is obeyed. The value of the velocity constant at $T = 293°$ K is $0{\cdot}010$ min.$^{-1}$.

In experiments where carbon monoxide and oxygen are admitted together to the deoxygenated surface, the adsorption of carbon monoxide and the reaction between $CO + O_2$ both occur. It is necessary to make allowance for the amount of gas adsorbed on the catalyst. It will be assumed that the adsorption processes are complete after 5 min. If V_0 is the volume of the carbon monoxide and oxygen present initially in ml. N.T.P. and $V_{\text{Ads.}}$ the volume of carbon monoxide adsorbed after 5 min., then the total volume at 5 min. is $V_0 - V_{\text{Ads.}}$. $V_0 = \frac{1062}{76} p_0$, where 1062 ml. is the volume of the apparatus and p_0 is the pressure in cm. At a time t, the total volume of gas present in the apparatus is $\frac{1062}{76} p$. Volume of gas removed by the reaction is $\frac{1062}{76} p_0 - V_{\text{Ads.}} - \frac{1062}{76} p$, and the volume of oxygen reacted is one third of this. The volume of oxygen present at any time t is $V^0_{O_2} - \frac{1}{3} \frac{1062}{76} (p_0 - p) + \frac{1}{3} V_{\text{Ads.}}$ and the partial pressure of the oxygen, $p_{O_2} = 0{\cdot}0716\, V^0_{O_2} - \frac{1}{3}(p_0 - p) + \frac{1}{3} 0{\cdot}0716 V_{\text{Ads.}}$. The derivation of the partial pressure of the oxygen requires a knowledge of $V_{\text{Ads.}}$, and if this knowledge is not available it is necessary to write the following logarithmic equation in the exponential form

$$\log_{10}\left\{0{\cdot}0716\, V^0_{O_2} - \tfrac{1}{3}(p_0 - p) + \frac{0{\cdot}0716}{3} V_{\text{Ads.}}\right\} = -\frac{kt}{3 \times 2{\cdot}303} + \log_{10} 0{\cdot}0716\, V^0_{O_2}.$$

Inserting the value found above for k, this becomes

$$0{\cdot}0716\, V^0_{O_2} - \tfrac{1}{3}(p_0 - p) + \frac{0{\cdot}0716}{3} V_{\text{Ads.}} = 10^{-0{\cdot}0015t} \times V^0_{O_2} \times 0{\cdot}0716,$$

or

$$p + (3 \times 0{\cdot}0716\, V^0_{O_2} - p_0 + 0{\cdot}0716 V_{\text{Ads.}}) = 3 \times 0{\cdot}0716\, V^0_{O_2} \times 10^{-0{\cdot}0015t}.$$

Hence a graph of p against $10^{-0{\cdot}0015t}$ should be a straight line with a gradient of $3 \times 0{\cdot}0716 \times V^0_{O_2}$. Graphs for these experiments are given in figure 5, and it will

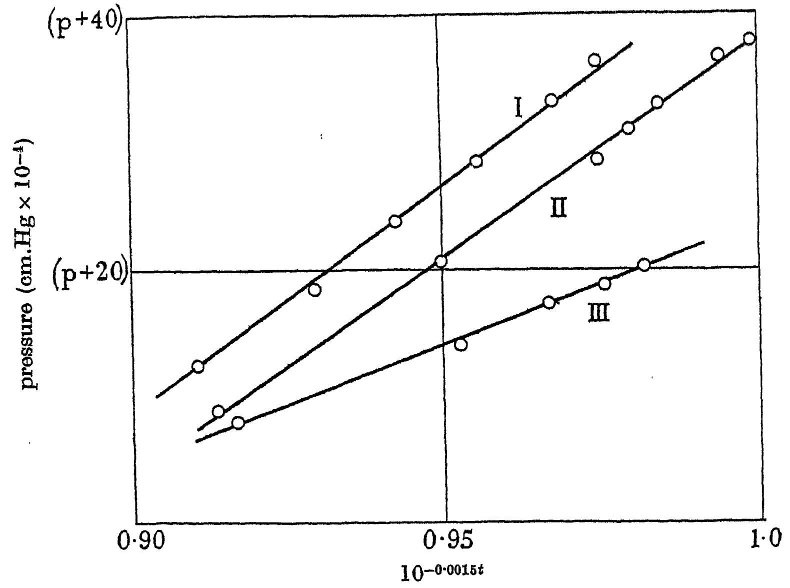


FIGURE 5. Reaction between carbon monoxide and oxygen.

be seen that the runs give good straight lines. The slopes should be proportional to the initial pressure of the oxygen, and this is shown to be true by table 3.

TABLE 3

curve	$p^0_{O_2}$ from measurement	$p^0_{O_2}$ from graphs
I	0·010	0·012
II	0·011	0·012
III	0·0054	0·006

The analysis has been carried further to deduce $V_{\text{Ads.}}$, and no inconsistency has been found. This treatment affords very considerable support to the views advanced above with regard to the kinetics of the reaction.

Heat of adsorption of mobile oxygen

The heat liberated during the adsorption of oxygen on a surface deoxygenated with carbon monoxide has been measured several times in an effort to obtain the heat of adsorption of mobile oxygen. The total amount of gas taken up is, however, very small, *c.* 0·015 ml., and the adsorption continues for a long time. That taken up immediately, which has an effect on the conductivity, is still smaller. The galvanometer deflexions are too small to give very accurate results. The heat was found to vary with time and increased from 35 to 75 kcal./mol. From the evidence yielded by conductivity measurements, the heat of adsorption of mobile oxygen should not exceed 30 to 35 kcal. It was clear, therefore, that in the above measurements some oxidation of the surface was occurring. The removal of carbon dioxide from the surface which occurs during the deoxygenation process leads to some reduction, and it is presumed that some of the oxygen is taken up in the restoration of the surface. It would, however, be preferable to prepare the surface by deoxygenation through evacuation at 200° C rather than with carbon monoxide. For accurate heats to be obtained, however, a much larger area of oxide would be required.

Adsorption of oxygen on a surface saturated with carbon monoxide

After adsorption of carbon monoxide, the catalyst is evacuated at room temperature for 1 hr. and oxygen is admitted. A heat of adsorption of 60 to 65 kcal./mol. oxygen was obtained. About 50 % of the adsorbed carbon monoxide can be removed under vacuum at room temperature as carbon dioxide (cf. Garner 1947).

CONDUCTIVITY MEASUREMENTS OF FILMS OF METAL AND OXIDE

Copper was evaporated from a coated tungsten wire in a vacuum of 10^{-7} to 10^{-8} cm. on to the inside surface of a Pyrex cylinder, and films of copper oxide prepared from the metal film by oxidation. The resistance of the films is measured between circular platinum electrodes, fused to the inside surface of the cylinder, using potentiometric and a.c. bridge methods, as described in the following paper.

The change in resistance was followed during oxidation at 200° C. This was carried through to cupric oxide and until a constant resistance was obtained. The initial period in the oxidation was characterized by a parabolic uptake of oxygen. The cupric oxide was reduced with hydrogen, and immediately on adding hydrogen there was an initial small increase in resistance. This was followed by an induction period, during which it is likely that copper nuclei are formed on the surface of the oxide. After 40 min. there was a rapid rise in the resistance, up to a maximum of several megohms, and the maximum was found to correspond stoichiometrically with the composition of cuprous oxide. After this maximum there was an equally rapid fall in resistance down to approximately 20 kΩ, when the resistance became subject to rapid fluctuations which lasted 20 to 30 min. After this period there was a continuous fall in resistance down to that of the initial resistance of pure copper after 30,000 min. (figure 6).

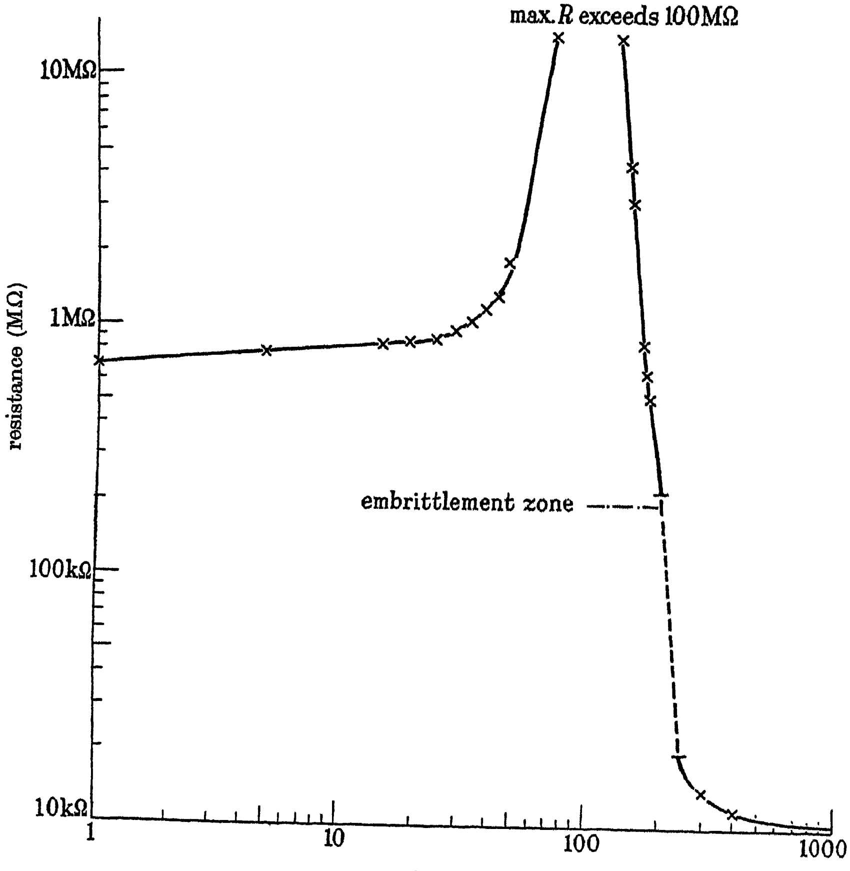


Figure 6. Reduction of copper oxide by hydrogen at 200° C. - - - Resistance readings highly erratic varying between 250 and 20 kΩ. After 30,000 min. R_{Final} became that of the initial copper film.

The region of erratic readings corresponds with the 'embrittlement' of the copper, and is probably due to the reduction of oxide inclusions by hydrogen diffusing through the metal. The escape of water along grain boundaries produces cracks in the film which, however, heal when the embrittlement process is completed. After embrittlement, the metal film has been ‘activated’.

Experiments were carried out to study the effects of the adsorption of gases on the electrical conductivity. The films were employed in the semi-conducting region between cuprous and cupric oxide, and it was found that over a wide range of compositions in this region the resistance was approximately constant.

Oxygen

In one series of experiments the initial resistance in a hard vacuum of a Cu_2O-CuO film was 650 kΩ. Admission of oxygen at 200° C at pressures ranging from 10^{-4} to 0·03 cm. caused an immediate fall in resistance to 70 kΩ, which was completed within the first 5 sec. (figure 7). The fall in resistance was independent of the oxygen pressure over the range of pressures studied. This was followed by

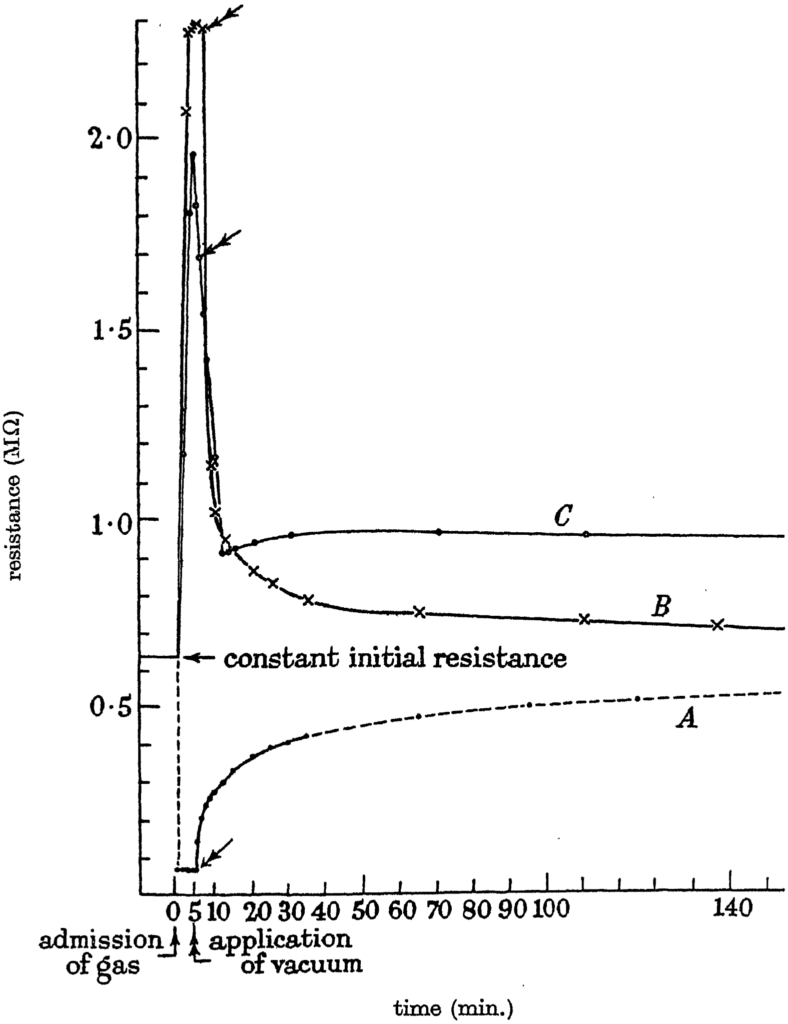


FIGURE 7. Adsorption of gases on cuprous oxide at 200° C. *A*, oxygen; *B*, hydrogen; *C*, carbon monoxide.

a slower change over the next few minutes, which was probably due to disappearance of the oxygen into the Cu_2O-CuO interface. On applying a vacuum of 10^{-7} to 10^{-8} cm., which was reached in 15 sec., the conductivity of the oxygenated oxide decreased as the square root of the time, $\Delta\lambda = kt^{-\frac{1}{2}}$ (figure 8), and returned to the initial value of 650 kΩ after 2500 min. This process could be repeated many times,

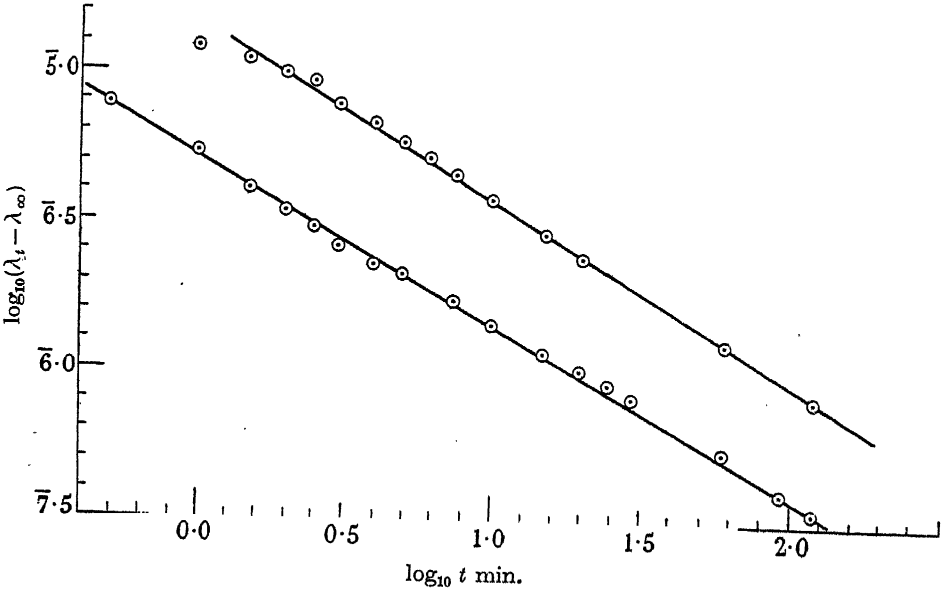


FIGURE 8. Desorption of oxygen (at 200° C).

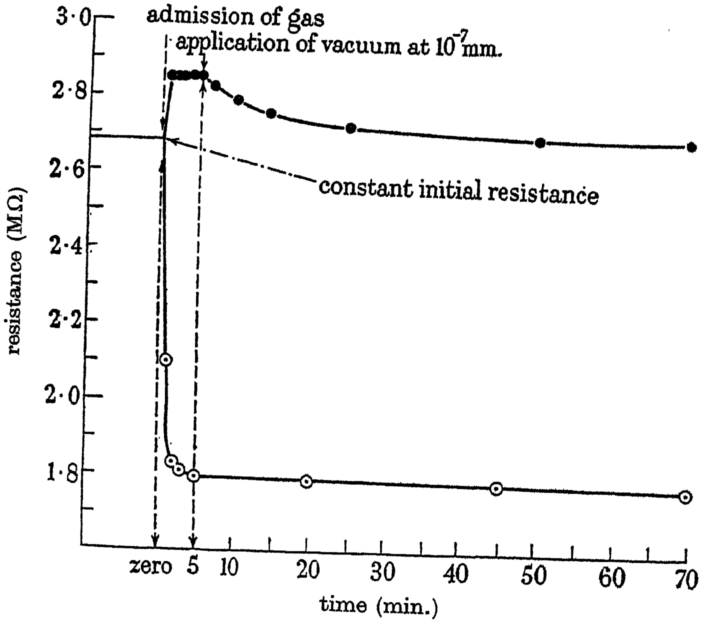


FIGURE 9. Adsorption of carbon monoxide and oxygen on copper oxide at 20° C. ⊙, oxygen; ●, carbon monoxide.

and the curves obtained fell closely on one another. There was retardation of the rate of change of resistance when the helium was admitted before the oxygen, both on the falling and rising curves, indicating that the diffusion of oxygen down cracks in the oxide was being hindered by the inert gas. This was to be expected, since the solid had been previously activated by embrittlement, which process has been shown to disrupt the lattice.

At lower temperatures, *c.* 100° C, the adsorption is still very rapid, but the change occurring on evacuation becomes much slower. At room temperature, although the adsorption was still complete within 5 min., there was no measurable increase in electrical conductivity on evacuation (figure 9), showing that the effects produced by adsorbed oxygen are relatively stable at room temperature. The effect of oxygen on an oxide surface activated by hydrogen is much greater than an oxide surface prepared directly from the metal.

Hydrogen

On the admission of hydrogen at 200° C (figure 7), the resistance increases very rapidly from 650 kΩ to 2·3 MΩ, and then becomes constant. After 5 min. exposure to hydrogen, a vacuum was applied and the resistance fell back to its initial value. The decrease in conductivity which occurs during the desorption obeys the equation $\Delta\lambda = kt^{-\frac{1}{2}}$.

Carbon monoxide

Carbon monoxide exerted a very similar effect (figure 7), except that in this case the resistance rose to a maximum and then fell spontaneously. On applying a vacuum, however, the fall was accelerated down to a final value of 1 MΩ, the resistance not returning to its initial value. There is evidence from the curves that three distinct processes are occurring, (1) an initial rise in resistance, (2) a rapid fall followed by (3) a small rise in resistance. After treatment with carbon monoxide at 200° C, followed by its removal on evacuation, the activity of the film was decreased. The fall in resistance with oxygen was very much smaller than was obtained with the original film. The normal activity could only be restored by reoxidizing and reducing with hydrogen, passing through the embrittlement period. The films, however, became unstable after treatment with carbon monoxide at 200° C, and tend to come away from the glass during the embrittlement process, so that but few experiments with carbon monoxide can be done with any one film. This effect was more marked with thick than with thin films.

The curves in figures 9 and 10 show the effects of carbon monoxide at room temperature on a Cu_2O-CuO surface which has not been activated with hydrogen, but prepared directly from the metal. The increase in resistance on adding carbon monoxide to a deoxygenated surface is relatively much smaller at room temperature than at 200° C, and the process is reversible, the resistance returning to the initial stationary value on evacuation (figure 9). The increase in resistance on adding carbon monoxide to an oxygenated surface is greater than on a deoxygenated surface (figure 10). There is also a decrease in the resistance on evacuation, but the fall is much less than the rise on the addition of carbon monoxide.

Similar results were obtained on an activated CuO-Cu_2O surface, but although in this case the conductivity changes on adding oxygen were greater, the effects of carbon monoxide additions were very similar to those shown in figure 10.

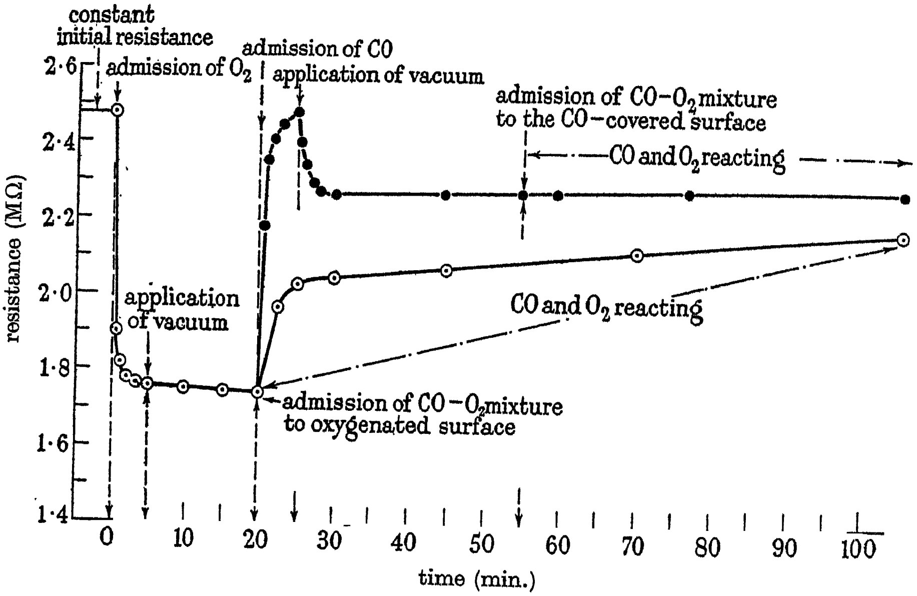


FIGURE 10. The reaction between carbon monoxide and oxygen on copper oxide at 20° C.

CO + *excess* O_2

If a mixture of carbon monoxide and oxygen was added to a surface after treatment with carbon monoxide, the catalytic reaction to give carbon dioxide took place without change in the surface conductivity (figure 10). On the other hand, when this mixture was added to an oxygenated surface, the resistance rose slowly to the same value obtained with a carbon monoxide-treated surface. It may be concluded from these results that in the steady state for the carbon monoxide-oxygen reaction, the surface is saturated with carbon monoxide. This confirms the results obtained in the adsorption calorimeter.

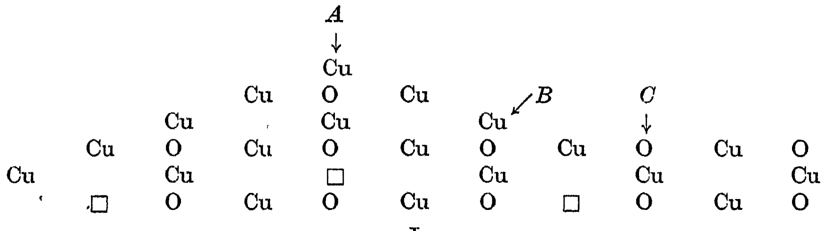
INTERPRETATION OF THE CHANGES IN RESISTANCE ON ADSORPTION OF GASES

Oxygen

The rise in conductivity on adsorption of oxygen may be due to the adsorption of oxygen as a negative ion, and in order to maintain electrical neutrality a positive hole, $Cu^+ \rightarrow Cu^{++}$ is created at the same time. The conductivity could then be due to the movement of electrons across an array of Cu^+ and Cu^{++} ions; that is, the superficial conductivity may be mainly electronic. From the fact that the adsorbed oxygen can be easily removed at 200° C it is likely that the heat of desorption does not exceed 30 to 35 kcal./g.mol., and this small heat is an indication that the gas is adsorbed as molecules. The measurements of the heat of adsorption of oxygen are not, however, conclusive in this respect.

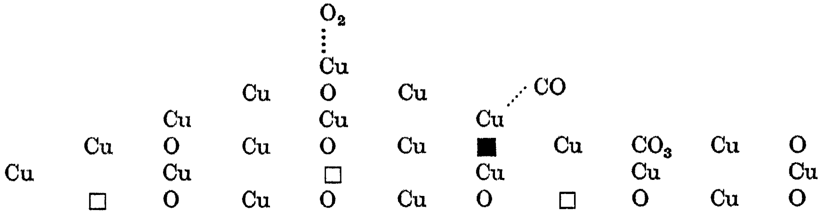
Since the conductivity in the presence of oxygen is permanently reduced after treatment of the surface by carbon monoxide at 200° C, it is concluded that the adsorption of the gases occurs on specially active areas of the surface, the activity of which is destroyed by the reaction. These active areas may consist of copper atoms exposed above the surface of the normal lattice.

A representation of such areas is given in diagram I, which shows, as a section across the crystal, the 100 plane of Cu_2O, □ representing vacant lattice sites left in position during the oxidation of the copper:



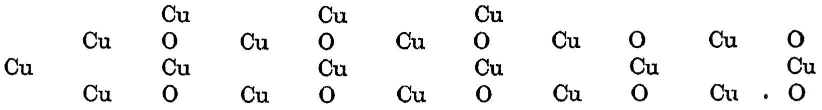
I

The active areas on which it is suggested that oxygen is adsorbed are shown at *A*. On such places it would be expected that oxygen would be loosely held. The destruction of the active areas by carbon monoxide could then occur by the adsorption of carbon monoxide at *B* and *C* as in diagram II. At *B* the CO is adsorbed reversibly, and at *C* as a carbonate ion. In the formation of this ion, a vacant oxygen lattice site ■ is produced which increases the instability of the surface.



II

On the desorption of the gases from the surface, the activity is partially destroyed as in diagram III, vacant lattice sites disappearing in the process:



III

Carbon monoxide

The mechanism of the adsorption of carbon monoxide can be illustrated by means of diagram II. On a *deoxygenated* surface at room temperature it is in part adsorbed reversibly as CO and in part as carbonate ion. Only the reversibly held carbon monoxide can be removed by evacuation at room temperature. The resistance changes on adding carbon monoxide are reversible at these temperatures (cf. figure 9), showing that reversibly adsorbed carbon monoxide immobilizes the conductance electrons of the solid. It also follows that the carbon monoxide-yielding stable carbonate ions does not materially affect the conductivity of the solid.

On the *oxygenated* surface at room temperature, it has been shown that the first step is the adsorption of carbon monoxide, followed by the removal of the mobile oxygen as carbon dioxide. The changes in resistance accompanying the addition of carbon monoxide to the oxygenated surface are shown in figure 10.

The effects of the addition of carbon monoxide at 200° C on the electrical resistance of the surface, deoxygenated by long evacuation, are shown in figure 9. Three stages are observed, and these are interpreted in the light of the above argument as due to the occurrence of three consecutive processes: (1) the adsorption of carbon monoxide on sites *B*, with an increase in resistance; (2) a reaction between mobile oxygen at *A* and carbon monoxide at *B* to give an unstable carbonate ion with a fall in resistance, and (3) evolution of carbon dioxide with a rise in resistance. At the conclusion of these processes the surface has been partially deactivated, as illustrated above.

Reaction between carbon monoxide and oxygen at room temperature

The experiments show that this reaction occurs on a surface saturated with carbon monoxide (experiments in adsorption calorimeter and figure 10). The reaction between oxygen and carbon monoxide is envisaged to occur on adjacent sites *A* and *B*. Figure 10 shows that sites *A* are not appreciably covered with oxygen during the reaction, so that these sites are mainly occupied by carbon monoxide, and hence the rate of reaction is proportional to the oxygen pressure.

DISCUSSION

The phenomena described in this paper are concerned with the interaction of oxygen, carbon monoxide and hydrogen with the copper-cuprous oxide-cupric oxide system. Only the first two gases have been studied in any detail and will be discussed here.

It has been demonstrated that oxygen is reversibly chemisorbed on a cuprous-cupric oxide surface in a form which is specially active chemically. In this state it will combine with carbon monoxide to give carbon dioxide at room temperature. This mobile form of oxygen is associated with a conductance electron, and its adsorption increases the conductivity of the oxide film. From the fact that the activity of the surface is very largely destroyed by treatment with carbon monoxide at 200° C, it is concluded that the adsorption of oxygen giving rise to the superficial conductivity occurs on atoms of copper exposed above the normal lattice. A mechanism for the deactivation has been given in the body of the paper.

At 200° C the rate of evaporation of chemisorbed oxygen is appreciable, but it does not appear to be entirely removed from the active areas even in a hard vacuum, because it is slowly replenished from the body of the lattice, possibly from the CuO-Cu_2O interface. In the presence of oxygen gas, saturation of the active points is reached at a relatively low pressure, below 10^{-4} cm., at 200° C. At room temperature, the chemisorbed oxygen is relatively immobile, and it is not evaporated from the surface at appreciable rates in hard vacuum.

Catalytic reaction at room temperature

The experimental results indicate that the reaction occurring between carbon monoxide and oxygen on activated copper oxide at room temperature is primarily a reaction between reversibly absorbed carbon monoxide, which decreases the electrical conductivity of the surface, and adsorbed oxygen atoms or molecules which increase the conductivity. The two gases are probably adsorbed on similar adjacent sites on the active areas. It is deduced that an unstable carbonate ion is formed which finally breaks down to give carbon dioxide. Carbon monoxide is much more strongly held than oxygen on the active areas; therefore the reaction is independent of the carbon monoxide pressure and proportional to the oxygen pressure.

During the reaction the surface is saturated by carbon monoxide, some of the gas being held reversibly as carbon monoxide, probably on exposed copper atoms, and some irreversibly as carbonate ion. The major part of the adsorbed carbon monoxide is held as carbonate ion on less active portions of the surface and does not appear to take any active part in the reaction at room temperature. It stabilizes the somewhat unstable surface by preventing reduction of cuprous oxide to metal, and limits the reaction to the active areas. On raising the temperature, however, the carbon dioxide held on the less active parts of the surface is evolved. At 200° C and above, the carbon monoxide-oxygen reaction therefore spreads to the less active parts of the surface, and mechanisms come into play whereby the carbon monoxide reacts directly with the normal oxygen ion of the lattice. The occurrence of this reaction gradually deactivates the surface since it leads to the diffusion of the exposed copper atoms into the body of the lattice, a process which is analogous to sintering.

[*Note added in proof*, 31 *March* 1949.] Professor Mott has pointed out that the expression for the change in conductivity occurring during the desorption of oxygen, (1) $\Delta\lambda = kt^{-\frac{1}{2}}$ (p. 308), can be interpreted as follows: the conductivity is proportional to n, the number of electrons in the conduction band, or the number of positive holes in the full band and from equation (1), $dn/dt = \text{const.}\ n^3$. Since the relationship between n and the number of impurity centres (adsorbed oxygen atoms) N is $n = \text{const.}\sqrt{N}$, it follows that $-dN/dt = \text{const.}\ N^2$. Thus the recovery after oxygenation is to be interpreted as due to a bimolecular process in which oxygen atoms, mobile over the oxide surface give oxygen molecules which then evaporate.

The authors wish to express their thanks to Imperial Chemical Industries Ltd. for a grant for the purchase of apparatus, and one of us (F. S. S.) his indebtedness to the Department of Scientific and Industrial Research for a maintenance grant, and to Monsanto Chemicals Ltd. for a Fellowship.

References

Constable, F. H. 1927 *Proc. Roy. Soc.* A, **115**, 570.
Constable, F. H. 1928 *Proc. Roy. Soc.* A, **117**, 376.
Garner, W. E. 1947 *J. Chem. Soc.* p. 1239.
Garner, W. E. & Veal, F. J. 1935 *J. Chem. Soc.* p. 1436.
Hasegawa, S. 1943 *Proc. Imp. Acad. Japan*, **19**, 393 (*Amer. Chem. Abstr.* 1947, 5777).
Hinshelwood, C. N. 1923 *Proc. Roy. Soc.* A, **102**, 318.
Jones, H. A. & Taylor, H. S. 1923 *J. Phys. Chem.* **27**, 623.
Juliard, A. 1932 *Bull. Soc. chim. Belg.* **41**, 65, 138, 234, 289 (*Chem. Zbl.* 1932, **2**, 2592).
Lamb, A. B., Bray, W. C. & Frazer, J. C. W. 1920 *J. industr. Engng Chem.* **12**, 213.
Lustman, B. 1942 *Trans. Electrochem. Soc.* **81**, 359.
Palmer, W. G. 1923 *Proc. Roy. Soc.* A, **103**, 444.
Pease, R. N. & Taylor, H. S. 1921 *J. Amer. Chem. Soc.* **43**, 2179.
Pease, R. N. & Taylor, H. S. 1922 *J. Amer. Chem. Soc.* **44**, 1637.
Ransley, C. E. 1939 *J. Inst. Met.* **65**, 147.
Rhines, F. N. & Anderson, W. A. 1941 *Trans. Amer. Inst. Min. (Metall.) Engrs*, **143**, 312.
Wagner, C. 1933 *Z. phys. Chem.* **21**, 25.
Wagner, C. & Grünewald, K. 1938 *Z. phys. Chem.* B, **40**, 455.
Wischin, A. 1939 *Proc. Roy. Soc.* A, **172**, 314.
Wright, C. R. A. & Luff, A. P. 1878 *J. Chem. Soc.* **33**, 1.
Wright, C. R. A., Luff, A. P. & Rennie, E. H. 1879 *J. Chem. Soc.* **35**, 475.

The application of semi-conductivity measurements in the study of catalysis on copper oxide

By T. J. Gray, *Department of Inorganic and Physical Chemistry, The University of Bristol*

(*Communicated by W. E. Garner, F.R.S.—Received* 6 *September* 1948 —*Read* 17 *March* 1949)

An apparatus has been constructed to measure the electrical conductivity of thin films of oxides in order to study the effects of adsorbed gases and reactions between them. This has been applied to the cuprous-cupric oxide system.

Introduction

The study of the semi-conductivity of copper oxide has revealed its dependence on the stoichiometric excess or defect of one component. Pfund discovered as early as 1916 that cuprous oxide showed semi-conductivity and photoconductivity, but it was not until 1930–2 that the part played by excess of oxygen in the conductivity of cuprous oxide was appreciated. Von Auwers & Kerschbaum (1930) found a reduction in the semi-conductivity of copper oxide plates when these were heated *in vacuo*, and in the following year Le Blanc & Sachse (1931), in a more systematic study, investigated the specific conductivity of compressed copper oxide powder with varying oxygen excess. They found a specific conductance *in vacuo* of 10^{-9} mho/cm. at 20° C, which changed to 10^{-6} mho/cm. when 0·1 %

oxygen was present. Dubar (1931) made similar observations with a slightly different technique, as also did von Waibel (1931). Jusé & Kurtschatow (1932) investigated the semi-conductivity of specimens obtained by the high-temperature oxidation of copper strip. They studied the effect of oxygen excess in the sample and the effect of an oxygen atmosphere, observing a very great increase in specific conductance on exposing evacuated samples to oxygen. They also found that the specific conductivity of cuprous oxide with but 0·6 % cupric oxide was very little different from that of pure cupric oxide.

The investigation was continued in the work of Dunwald & Wagner (1933), which was largely confined to high temperatures, and this enabled Wagner (1933) to develop a theoretical relationship between the semi-conductivity and oxygen excess. This was further advanced by Wagner & Hamman (1938). In a further contribution Dubar (1938) differentiated between bulk and superficial conductivity. The adsorptions of oxygen, moist and dry air, and nitrogen were investigated, using thin strips of cuprous oxide, and a study was also made of the effect of superficial chemical treatment. He found that adsorbed gases had greater influence on the superficial conductivity of the oxide than on the bulk conductivity.

In general, the experimental methods have been made on bulk oxides usually prepared by the high-temperature oxidation of copper strip, although in some cases compressed powders have been used. The normal method of measuring resistance has been by measurements of current with a secondary cell and galvanometer, but occasionally bridge measurements have been made. The ohmic nature of the resistance has usually been demonstrated over ranges varying from 0 to 10 V/cm. In the case of Jusé & Kurtschakow's researches, quoted by Dubar, the range of measurements extended between 0 and 7000 V/cm.

The method of investigation now to be described is based on the measurement of the change in conductivity of an oxide due to the adsorption or desorption of various gases. The stationary state of conductivity obtained by prolonged heating in a vacuum of 10^{-7} mm. at a suitably selected temperature is taken as a datum line, and this is normally established between successive adsorptions. In the interpretation of the results it has been assumed that for instantaneous or very rapid changes of conductivity occurring on the adsorption of a gas, the incremental change in conductivity is directly proportional to the amount of the adsorbed gas. The present application employs the property of semi-conductivity to obtain a qualitative picture of the processes of adsorption and catalytic reaction while these processes were being quantitatively followed by alternative methods.

General considerations concerning the adoption and development of a method for the investigation of the adsorption of gases based on changes in conductivity indicate, as a first essential, the necessity for the preparation of the catalyst as a film of uniform thickness between the conduction electrodes. The most satisfactory method appears to be the condensation of metal vapour on a surface between the electrodes and the conversion of the metal film to the oxide. In order to prevent contamination of the oxide with mercury and grease, it is advisable to use greaseless high-vacuum systems. The properties of the metallic film are seriously affected by very small amounts of mercury. Suitable provision must be made for the introduction of gases, the measurement of pressures and very high-speed pumping for desorption. The electrical considerations demand great versatility from the measuring instruments; to have a range of from 0·01 Ω to 1000 MΩ (to be measured at both d.c. and a.c.) and to be capable of very rapid operation. These requirements have been met by the employment of an electronic potentiometer and an a.c. bridge. Since the temperature coefficients of conductivity encountered in semi-conductors are very large, precise temperature control is essential, and this is achieved by the use of an electronic thermal regulator operated from an a.c. platinum resistance thermometer.

Experimental

Several forms of reaction vessel have been developed, the form in current use being described in figure 1. The conduction electrodes of pure platinum are fused into the inner surface of a hard glass cylinder, four electrodes being provided uniformly spaced and accurately calibrated. Conduction leads of heavy-gauge platinum wire are connected through the cylinder on to the electrodes and are led out of the top of the reaction tube through intermediate seals. The evaporation filament is supported separately, and the plated tungsten wire is mounted coaxially with the cylinder bearing the electrodes and is spring-loaded to maintain its position when heated. Special precautions are taken during the electroplating of the tungsten wire to maintain the uniformity of deposit and to ensure the maximum cohesion.

The reaction tube is a very close fit in the bore of the furnace, and is joined to the vacuum system with a wide-bore side tube 4 cm. in diameter at the upper end. The tube is subjected to normal cleaning technique, and in addition a high-frequency discharge in low-pressure inert gas is operated during the evacuation and baking-out processes. This has been found to be essential to produce coherent evaporated films.

The reaction tube is protected by two liquid-oxygen traps from the mercury vapour present in the main reaction system, which is illustrated in figure 2. Large-bore mercury cut-offs, diameter 2 cm., are used throughout. A calibrated expansion bulb and McLeod and Pirani gauges are included in the reaction system for adsorption and pressure measurements. The gases are introduced from a gas burette protected by a cut-off, or by silver or palladium thimbles, as appropriate. The pumping system, which is not illustrated, comprises two large-capacity glass mercury-vapour pumps working in tandem and capable of reducing the reaction system from 5 to 10^{-7} mm. in less than 15 sec.

The furnace unit is illustrated in figure 3. The furnace block is of stainless steel 30 cm. in length, 12 cm. in diameter and with an axial bore of 4 cm. Six exponentially wound heaters are hexagonally disposed on a radius of 3·5 cm., and these, together with a compensating end heater, achieve uniformity of temperature throughout the length of the furnace. The controlling platinum resistance thermometer is located in a hole adjacent to the main bore. The furnace block has substantial thermal insulation and can be raised and lowered by a counterpoise system on vertical steel bars. Manual control is used to maintain the furnace approximately 2½° C below the working temperature, while electronic control in the form of an a.c. bridge operating from the platinum resistance thermometer gives a precision of better than 0·01° C at all temperatures.

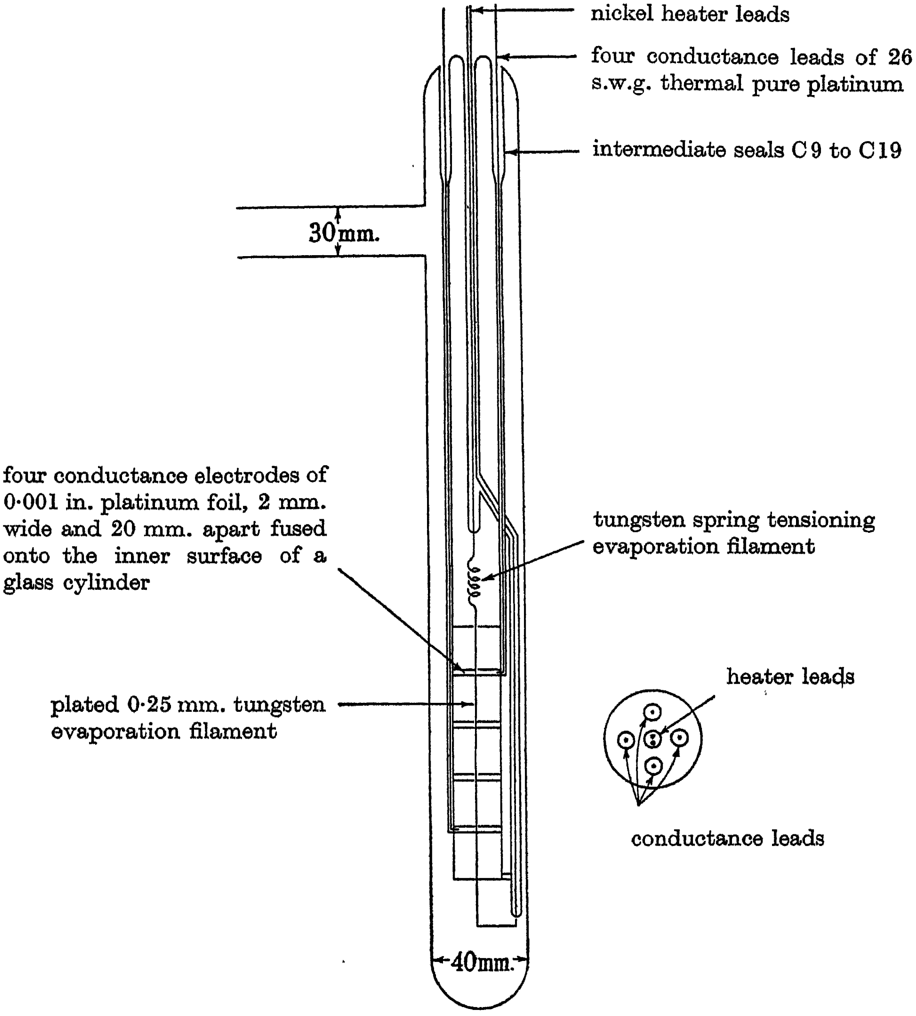

nickel heater leads
four conductance leads of 26 s.w.g. thermal pure platinum
intermediate seals C 9 to C 19
30 mm.
four conductance electrodes of 0·001 in. platinum foil, 2 mm. wide and 20 mm. apart fused onto the inner surface of a glass cylinder
tungsten spring tensioning evaporation filament
heater leads
plated 0·25 mm. tungsten evaporation filament
conductance leads
40mm.

FIGURE 1. Reaction tube 'B'.

A precision Kohler potentiometer and standard resistors are employed in a conventional circuit for measurements of d.c. resistance. A.c. measurements are made with a normal a.c. bridge in which the ratio arms are replaced by alternative potentiometers of 2000 and 200,000 Ω for the low- and high-resistance ranges respectively. These are large-diameter rotary potentiometers, and by limiting the ratio to 3·5 to 1 in either direction an accuracy of ½ % up to 35 MΩ and 2 % from 35 to 350 MΩ has been obtained. Internal standards are employed with provision for the use of external standards. The detector is a cathode follower and amplifier feeding a tuning indicator or phones.

Provision is also made for the investigation of the adsorption of gases in the presence of d.c. fields and a two-stage degenerative voltage regulated power supply has been developed, adjustable up to 500 V with a regulation of 1 part in 25,000.

Outline of method

The apparatus has been employed to study adsorption on a copper oxide film. The plating of a tungsten wire 0·25 mm. in diameter with a uniform coherent film of copper is performed by normal plating technique using a cylindrical anode with the tungsten wire at the axis and with air agitation of the solution. No difficulty is experienced in plating tungsten providing the wire is carefully cleaned after removal of the oxide by a.c. electrolysis in 10% caustic soda. The uniformly plated wire is then spot-welded on to the evaporation electrodes of the reaction tube and carefully tensioned and alined. The apparatus is evacuated and thoroughly degassed, after which the metal is evaporated between the conduction electrodes by controlled heating of the filament. Very slow evaporation rates are used, and the simultaneous appearance of the first transmission colour over the whole surface is taken as a preliminary indication of a uniform film. Estimations of the thickness of the film are obtained in a variety of ways. The electrical resistance is taken as a first method, and successive films are matched by evaporating to a definite resistance between electrodes. Other estimates are obtained from the amount of metal evaporated from the filament and the amount on the cylinder when the apparatus is eventually cut down. These two estimates are made by normal analytical methods. Film thicknesses up to 2000 Å are used, with the average thickness being about 1200 Å. Films thicker than this tend to strip under subsequent treatment.

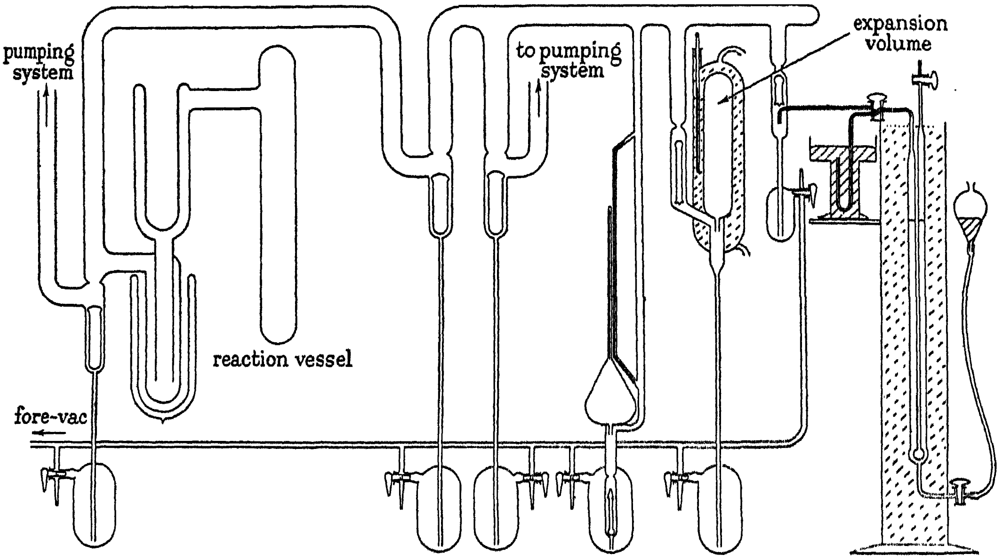


FIGURE 2. Reaction system.

The oxidation of the film is carried out with pure oxygen. The oxidation process is followed by pressure and resistance measurements. In some cases the oxidation is stopped at a predetermined point, whereas in other cases the film is completely oxidized, then reduced with hydrogen and reoxidized to the appropriate condition. The latter treatment activates the film of copper and increases the changes in conductivity obtained on adsorbing gases on the surface. The resistance is followed by both d.c. and a.c. measurements, and these afford a check on the uniformity of the film, since non-uniform films act as barrier cells and give varying values for the d.c., reversed d.c. and a.c. resistances.

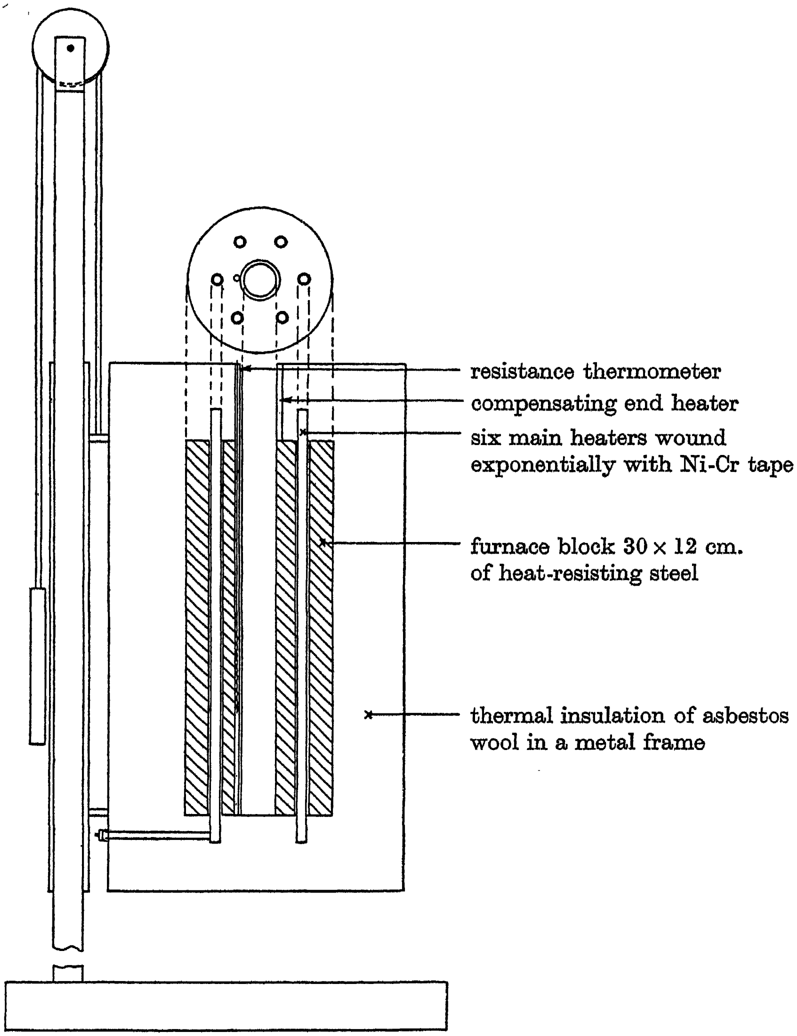

resistance thermometer
compensating end heater
six main heaters wound exponentially with Ni-Cr tape
furnace block 30 × 12 cm. of heat-resisting steel
thermal insulation of asbestos wool in a metal frame

FIGURE 3. Reaction furnace.

The copper oxide film in the semi-conducting state, which lies stoichiometrically between CuO and Cu_2O, is subjected to prolonged evacuation at 200° C at 10^{-7} mm. This establishes a constant stationary state, which is used as a datum line. The temperature coefficient of this value can be determined under conditions of hard vacuum. The adsorption of gases is now carried out at selected temperatures, the film being normally returned to its reference condition between adsorptions by prolonged evacuation at the reference temperature. This occurs readily in the case of oxygen and hydrogen, but carbon monoxide at the higher temperatures causes some deactivation of the film. Temperature coefficients for the change in conductivity on adsorption of various gases can be determined and the rates of desorption obtained against a vacuum of 10^{-7} mm. The amounts of gas adsorbed and changes in pressure are recorded.

Owing to technical difficulties the measurement of amounts of adsorption was introduced only towards the end of the series of experiments on copper oxide and relatively few values were obtained. In outline, the method adopted for these measurements entails the calibration of the apparent volume of the reaction system, using an inert gas for various temperatures of the reaction tube and for pressures ranging from 10^{-7} to 5 mm. Subsequently, by using calibrated expansion volumes (in addition to the normal gas burette) adsorption measurements can be made to 0·0001 ml. with satisfactory accuracy. It has not been possible to determine the effective area and roughness factor, the normal Brunauer, Emmett & Teller (1938) method being impracticable on such small superficial areas. The roughness factor of about 10 times has, however, been estimated in the preceding paper by interference methods for an oxide film formed on metallic copper.

Reactions between gases are investigated in a similar manner to simple adsorptions, using the characteristic increments obtained from the adsorption of the components as reference levels. This method is described in the preceding paper by Garner, Gray & Stone in relation to the reaction between carbon monoxide and oxygen over copper oxide catalyst, where it is possible to show that the reaction occurs on a surface saturated with carbon monoxide.

REFERENCES

Anderson, J. S. & Morton, M. C. 1945 *Proc. Roy. Soc.* A, **184**, 83.
Auwers, A. von & Kerschbaum, H. 1930 *Ann. Phys., Lpz.*, **7**, 129.
Brunauer, S. & Emmett, P. H. 1937 *J. Amer. Chem. Soc.* **59**, 1533, 2682.
Brunauer, S., Emmett, P. H. & Teller, E. 1938 *J. Amer. Chem. Soc.* **60**, 309.
Dubar, L. 1931 *C.R. Acad. Sci., Paris*, **192**, 341.
Dubar, L. 1938 *Ann. Phys., Paris*, **9**, 5.
Dunwald, H. & Wagner, C. 1933 *Z. phys. Chem.* **22**, 212.
Jusé, W. & Kurtschatow, I. V. 1932 *Phys. Z. Sowjet.* **2**, 453.
Le Blanc, M. & Sachse, H. 1931 *Ann. Phys., Lpz.*, **11**, 727.
Pfund, A. H. 1916 *Phys. Rev.* **7**, 289.
Wagner, C. 1933 *Z. phys. Chem.* B, **22**, 181.
Wagner, C. & Hamman, H. 1938 *Z. phys. Chem.* B, **40**, 197.
Waibel, F. 1931 *Wiss. Veröff. SiemensKonz.* **10**, 657.

The constitution of the aluminium-rich aluminium-cobalt-copper alloys, with special reference to the role of transitional metals in alloy formation

By P. C. L. Pfeil and G. V. Raynor

(*Communicated by A. J. Bradley, F.R.S.—Received* 4 *September* 1948—*Revised* 16 *December* 1948)

[Plate 14]

An experimental investigation of aluminium-rich aluminium-cobalt-copper alloys has shown that, as in the alloys of aluminium and copper with nickel and iron respectively, a ternary compound, in addition to the phases Co_2Al_9 and $CuAl_2$, enters into equilibrium with the primary solid solution. Isothermal sections of the ternary model have been established at 530° and 500° C; the field in which the solid solution, α, and the ternary compound, T(CoCu), are in equilibrium is very narrow, while the $(\alpha + Co_2Al_9)$, $(\alpha + Co_2Al_9 + T(CoCu))$ and $(\alpha + T(CoCu) + CuAl_2)$ phase fields are relatively extensive. The presence of T(CoCu), and its formation peritectically from Co_2Al_9, have been confirmed by further experiments on solid and semi-liquid alloys and, from the examination of slowly cooled alloys, the appropriate fields of primary crystallization have been determined. The composition range in which T(CoCu) separates as primary crystals is restricted, and, for this and other reasons, pure samples of T(CoCu) cannot be obtained for analysis. Extrapolation of the accurately established $(\alpha + T(CoCu) + CuAl_2)/(\alpha + T(CoCu))$ and $(\alpha + T(CoCu) + CuAl_2)/(T(CoCu) + CuAl_2)$ boundaries, however, shows that the homogeneity range of T(CoCu) includes the composition $Co_2Cu_5Al_{16}$. The solubility of copper in Co_2Al_9 does not exceed 1·44 % at 570° C.

In discussion, it is shown that the ternary compounds $NiCu_3Al_6$ and $FeCu_2Al_7$ occur at the same electron : atom ratio, according to the authors' theory of the role of transitional elements in alloy formation. The compound T(CoCu) forms a third member of the same series and is probably of the ideal composition $Co_2Cu_5Al_{13}$. The results support the hypothesis of absorption of electrons by transitional metal atoms present in aluminium-rich alloys, and also that the occurrence of ternary compounds is influenced to a marked degree by the electron : atom ratio.

1. Introduction

In recent papers (Raynor 1944*a*, 1945; Raynor & Little 1945; Raynor & Pfeil 1946–7*a*, *b*; Raynor & Wakeman 1947; and Raynor & Waldron 1948), the constitutions of a number of aluminium-rich ternary systems, containing, as solutes, one or two transitional metals of the first long period, have been described. This work was undertaken in order to obtain information, by systematic investigation, with regard to the role of transitional metals in this type of alloy. The results suggest very strongly that, when alloyed with a metal of relatively high valency such as aluminium, transitional metal atoms absorb electrons from the structure as a whole in such a manner as to fill the vacancies in their atomic orbitals. The importance of the electron : atom ratio as a factor in the formation of binary and ternary intermediate phases with aluminium and the transitional metals has also been established.

According to the theory of the electronic structure of transitional metals proposed by Pauling (1938), electrons derived from both the 3*d* and 4*s* levels of the free atom are to be regarded as concerned in the cohesion in the solid state. The electronic structure suggested is such that, for the metals chromium, manganese, iron, cobalt and nickel, 5·78 electrons per atom occupy 'bonding orbitals', and are concerned only in cohesion, while the remainder occupy 'atomic orbitals' which can contain a maximum of 4·88 electrons/atom. Thus chromium, manganese and iron have respectively 0·22, 1·22 and 2·22 electrons/atom in atomic orbitals; in each atom these electrons are 'unpaired', that is, they have the same value of the spin quantum number. For iron, the 2·22 unpaired atomic orbital electrons account for the saturation magnetic moment, at absolute zero, of 2·22 Bohr magnetons/atom. On passing to cobalt, the addition of one more electron/atom leads to the pairing of a certain number of electrons with electrons of opposite spin quantum number, since there is room in the atomic orbitals only for 2·44 unpaired electrons. As a result, cobalt possesses 1·51 paired electrons and 1·71 unpaired electrons/atom in atomic orbitals, while nickel, with one electron more, has 3·61 paired electrons and 0·61 unpaired electrons/atom in these orbitals. This conception implies the presence of vacancies in the atomic orbitals of the transitional metals of the first long period to the following extents per atom:

chromium	manganese	iron	cobalt	nickel
4·66	3·66	2·66	1·71	0·61

If it be assumed that, in aluminium-rich alloys, these vacancies may be filled by the absorption of electrons from the structure as a whole, several interesting analogies between different alloy systems may be traced. Thus, the systems aluminium-manganese-copper, aluminium-manganese-nickel (Raynor 1944*a*), and aluminium-manganese-zinc (Raynor & Wakeman 1947) each contain a ternary compound which enters into equilibrium with the primary aluminium-rich solid solution; in each case the compound is based on a ratio of four aluminium atoms to one solute atom, and, if manganese and nickel are considered to absorb respectively 3·66 and 0·61 electrons/atom, has an electron : atom ratio of 1·85. It has also been shown that the binary compound Co_2Al_9 and the ternary compound $FeNiAl_9$, the isomorphism of which may be accounted for by the fact that both phases have closely similar electron: atom ratios according to the present theory, will both dissolve nickel to the same electron : atom ratio of 2·285 (Raynor & Pfeil 1946–7*b*). On the basis of the theory, almost quantitative predictions of the forms of the equilibrium diagrams for the aluminium-rich aluminium-iron-cobalt and aluminium-cobalt-nickel alloys have been made (Raynor & Pfeil 1946–7*a*, *b*).

Although experimental evidence at present available suggests that the number of electrons absorbed per atom of a transitional metal is very similar to the number of atomic orbital vacancies, the physical mechanism by which absorption of this number of electrons occurs is not clear. Using the alternative band theory model, it might be expected that nickel, cobalt and iron would complete their 3*d* shells by the acceptance of 0, 1 and 2 electrons respectively. In the present paper, the non-integral values have been retained for consistency with previous work.

In the course of this work, it was noted that, in the systems aluminium-iron-copper and aluminium-nickel-copper, ternary compounds are formed which enter into equilibrium with the aluminium-rich solid solution. According to the results of Bradley & Goldschmidt (1939), and of Bradley & Lipson (1938), these compounds may be represented respectively as $FeCu_2Al_7$ and $NiCu_3Al_6$. Assuming that aluminium and copper respectively contribute three electrons/atom and one electron/atom to the structure, while iron and nickel respectively absorb 2·66 and 0·61 electrons/atom, the electron : atom ratios of $FeCu_2Al_7$ and $NiCu_3Al_6$ are 2·034 and 2·039. It appears probable, therefore, that the formation of these two compounds is dependent on the establishment of an electron : atom ratio of approximately 2·04. Assuming that nickel and iron absorb respectively 0 and 2 electrons/atom, the appropriate ratio is 2·1. If these views are correct, it would be expected that a ternary compound with a similar electron : atom ratio would occur in the aluminium-rich aluminium-cobalt-copper alloys. Since no information with regard to these alloys was available in the literature, an experimental investigation was undertaken in order to test this suggestion.

2. THE BINARY SYSTEMS

Figure 1 shows the aluminium-rich portion of the aluminium-cobalt equilibrium diagram according to Fink & Freche (1932). The solid solubility of cobalt in aluminium is negligibly small, and does not exceed 0·02 % at any temperature. At 1 % cobalt and a temperature of 657° C, the aluminium-rich solid solution and the compound Co_2Al_9 separate from the liquid as a eutectic mixture. The intermetallic compound, although based essentially on a ratio of two cobalt atoms to nine aluminium atoms, contains slightly more cobalt than that corresponding to the ideal formula (Raynor & Pfeil 1946–7 *b*).

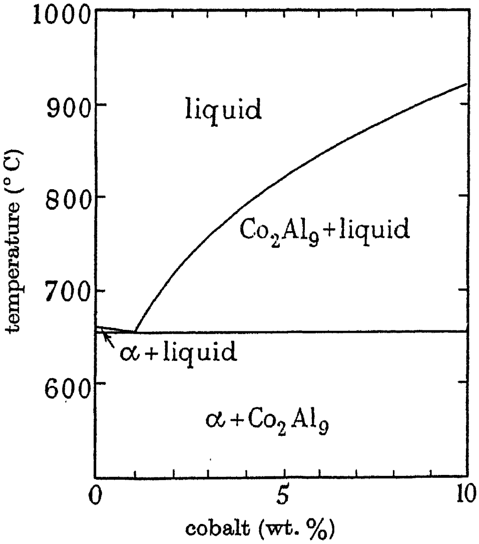


FIGURE 1. Equilibrium diagram for aluminium-cobalt system.

Existing data on the equilibrium diagram of the aluminium-copper alloys have been summarized by Raynor (1944 *b*), and the aluminium-rich region of the diagram is reproduced here as figure 2. The solid solubility of copper in aluminium extends up to 5·7 % copper at 548° C. At this temperature, and at a composition of 33 % copper, a eutectic is formed between the aluminium-rich solid solution and the compound $CuAl_2$.

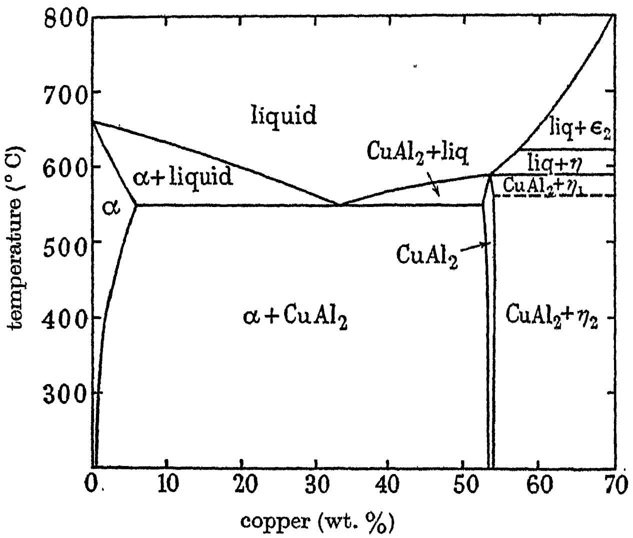


Figure 2. Equilibrium diagram for aluminium-copper system.

3. Materials used

Alloys were prepared from super-pure aluminium, obtained from the British Aluminium Company, Ltd., and from binary aluminium-cobalt and aluminium-copper master alloys. The aluminium-cobalt alloy was prepared from pure materials and presented by the British Aluminium Company Ltd., to whom grateful acknowledgement is made. The aluminium-copper alloy was prepared by the authors. The compositions of the master alloys are given in table 1.

Table 1

	Al-Co A	Al-Co B	Al-Cu A	Al-Cu B
cobalt (%)	4·65	11·07	—	—
copper (%)	0·025	—	34·0	61·4
silicon (%)	0·020	0·05	0·02	0·02
iron (%)	0·020	0·04	0·03	0·03
aluminium (%)	remainder	remainder	remainder	remainder

4. Experimental methods

Alloys for annealing experiments were prepared in quantities of 20 to 25 g. by melting together the appropriate quantities of the master alloys and aluminium in alumina-lined crucibles with vigorous stirring. In preliminary work, the alloys were cast into a heavy bronze mould. Owing, however, to the difficulty of attaining equilibrium on subsequent annealing, all the alloys on which the conclusions of this paper are based were cast in the special mould described by Raynor & Wakeman

(1948). In this apparatus, the alloy is poured into a thin-walled copper tube surrounded by a stirred freezing mixture, and very fine chill-cast microstructures are obtained.

Annealing was carried out in tubular resistance furnaces, automatically controlled to within ± 1° C of the required temperature. Specimens were enclosed in evacuated hard-glass tubes, and possible reaction between the metal and the glass was prevented by the interposition of a thin alumina sheath. Annealing treatments were concluded by quenching the specimens in cold water; freshly cut surfaces were prepared for micrographic examination by the usual methods.

Alloys for slow-cooling experiments were prepared in quantities of 60 g. After thorough stirring, the melts were cooled at 1·5° C/min., using a Foster Programme Controller. When cold, the ingots were sectioned longitudinally and examined micrographically. In certain cases, the primary crystals present were extracted electrolytically as previously described, and, after careful hand-sorting, submitted for analysis.

In all cases, critical alloys resulting from the annealing work were analyzed; alloys for which analyses were carried out, using the actual specimens examined micrographically, are distinguished in the diagrams contained in this paper by black symbols.

5. MICROGRAPHY OF THE ALLOYS

(*a*) *Alloys rich in aluminium*

In addition to the aluminium-rich solid solution, three metallographic constituents were recognized in annealed specimens:

(i) Co_2Al_9, which was grey in colour and showed relief effects.

(ii) $CuAl_2$, which was white with a pink tinge, and in low relief.

(iii) A ternary phase, denoted as T(CoCu) in this paper, which was a lighter grey, and in lower relief, than Co_2Al_9.

$CuAl_2$ could be distinguished from the other two phases in the unetched condition. Several etching reagents gave a satisfactory distinction between Co_2Al_9 and T(CoCu), the most useful being an aqueous solution containing 5 % sodium hydroxide and 5 % sodium carbonate. An immersion of 30 sec. stained Co_2Al_9 brown, while the ternary compound remained almost unaffected. In general, however, all critical alloys were examined using additional reagents for confirmation. Typical microstructures are reproduced in figures 3, 4, 5 and 6, plate 14.

(*b*) *Alloys containing* 48 *to* 55 % *copper*

The metallographic constituents recognized in annealed alloys rich in copper were:

(i) $CuAl_2$.

(ii) The aluminium-rich solid solution, appearing as small yellow areas, in lower relief than the matrix of $CuAl_2$.

(iii) T(CoCu), which retained the same etching characteristics as in the aluminium-rich alloys.

(iv) A grey phase, darker than T(CoCu), denoted as W in this paper. The etching characteristics of this phase were quite distinct from those of T(CoCu) and of Co_2Al_9.

(c) *Slowly cooled alloys*

In these alloys, the phases Co_2Al_9, T(CoCu) and $CuAl_2$ were recognized without difficulty, in addition to the aluminium-rich matrix. Two other phases were found in alloys rich in copper and cobalt. One of these was identified, from its etching characteristics, as the W phase; the second, a light grey phase resistant to attack by etching, has not been identified. Further reference is made to these phases below.

6. NOMENCLATURE

To facilitate reference to individual alloys in the following sections, alloys are identified by their compositions. Thus the term 'alloy 1·5/5·5' refers to an alloy containing 1·5 wt. % cobalt and 5·5 wt. % copper.

7. THE CONSTITUTION IN THE SOLID STATE

Initial experiments, in which alloys were annealed for 4 weeks, showed that equilibrium was only attained very slowly; several alloys contained the four phases α, Co_2Al_9, T(CoCu) and $CuAl_2$, where α, as elsewhere in this paper, denotes the aluminium-rich solid solution. In all later experiments, therefore, the alloys were cold-worked by hammering and squeezing in a vice before annealing, and between successive annealing periods, in order to hasten the attainment of equilibrium.

(a) *The* 530° C *isothermal—aluminium-rich alloys*

In general, alloys were examined after annealing for 4 weeks at 530° C, and again after a further annealing period of either 2 or 4 weeks at the same temperature. The general effect of this treatment was to bring alloys containing four phases after 4 weeks' annealing nearer to a normal three-phase structure. The amount of T(CoCu) present increased at the expense of either the $CuAl_2$ or the Co_2Al_9. All alloys in which any change in microstructure had occurred were again annealed, and re-examined at fortnightly intervals until a total annealing time of 12 weeks had been given, at which stage the experiments were discontinued. After 12 weeks, only three of the forty-eight aluminium-rich alloys contained traces of a fourth phase; these alloys are referred to in detail below.

The results of this work are summarized in figure 7, in which all the points plotted refer to alloys annealed for 12 weeks, or until no change in microstructure was obtained on successive annealing for a shorter total period. The $(\alpha + CuAl_2)$ and $(\alpha + T(CoCu))$ phase fields are very narrow, while the three-phase fields and the $(\alpha + Co_2Al_9)$ phase field are relatively extensive. The form of the $(\alpha + Co_2Al_9)$ field indicates that, at 530° C, the solubility of copper in Co_2Al_9 is not large, and probably of the order of 1 %.

The changes by which equilibrium is attained in alloys having compositions in the region of the $(\alpha + T(CoCu))$ phase field are of interest. The alloy 1·50/10·10 may be considered as an example. After annealing for 4 weeks, the α, Co_2Al_9 and $CuAl_2$ phases were present, together with a small but easily recognizable amount of T(CoCu). After 8 weeks the amount of Co_2Al_9 had decreased considerably, while the amount of T(CoCu) had increased correspondingly; the amount of $CuAl_2$ was relatively little changed. After a further fortnight, only traces of Co_2Al_9 remained, while after 12 weeks none was visible except as occasional minute inclusions in some of the larger T(CoCu) particles. It was concluded that the equilibrium structure of this alloy at 530° C was ($\alpha + T(CoCu) + CuAl_2$), and that annealing periods of prohibitive length would be necessary to remove the minute Co_2Al_9 inclusions.

Alloys 1·32/7·63 and 1·00/6·82 retained small amounts of $CuAl_2$ in structures which were predominantly ($\alpha + Co_2Al_9 + T(CoCu)$). Since these alloys were not critical, no annealing in excess of 12 weeks was given, and no reliance has been placed on them in figure 7.

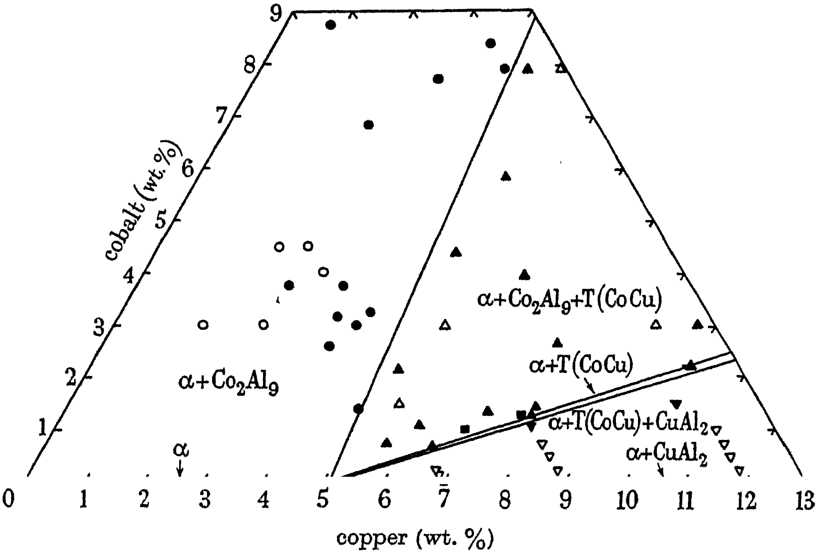


FIGURE 7. The 530° C isothermal; aluminium-rich alloys.

○ ● $\alpha + Co_2Al_9$. ▽▼ $\alpha + T(CoCu) + CuAl_2$. ■ 4-phase alloys.
△▲ $\alpha + Co_2Al_9 + T(CoCu)$. □ $\alpha + CuAl_2$.

(*b*) *The* 530° C *isothermal—alloys containing* 48 *to* 55 % *copper*

Eighteen alloys containing 48 to 55 % copper and up to 1·7 % cobalt were annealed for 3 weeks at 530° C, and after examination, the critical alloys were reannealed for a further 3 weeks. No changes in microstructure were observed; it is therefore probable that in this composition range equilibrium had been attained relatively rapidly. The results are plotted in figure 8. No attempt was made to define accurately the

$$(CuAl_2 + T(CoCu))/(CuAl_2 + T(CoCu) + W)$$

and

$$(CuAl_2 + W)/(CuAl_2 + T(CoCu) + W)$$

phase boundaries. The

$$(\alpha + T(CoCu) + CuAl_2)/(CuAl_2 + T(CoCu))$$

boundary is, however, accurately established, and extrapolates exactly to the aluminium-rich limit of $CuAl_2$ at 530° C as determined by Stockdale. The microstructures of the alloys upon which this boundary is based were easy to interpret, and very small traces of the α phase could be recognized without difficulty. The relative amounts of the α, T(CoCu) and $CuAl_2$ phases in alloy 0·21/51·90 indicated that the solubility of cobalt in $CuAl_2$ is restricted, and probably of the order of 0·1 % cobalt.

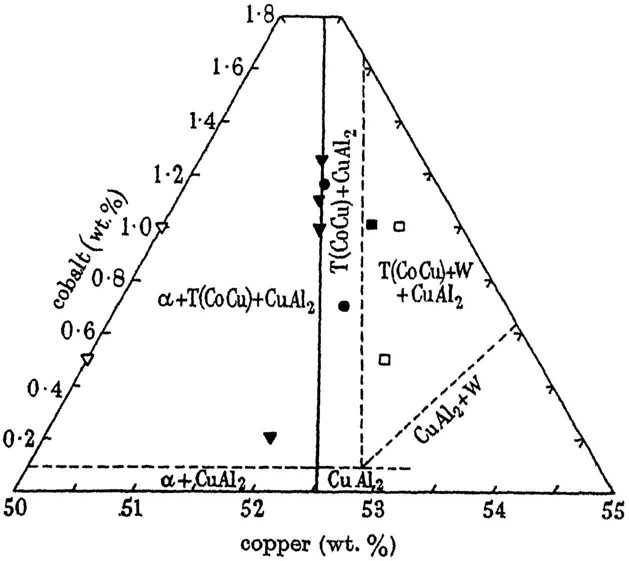


FIGURE 8. The 530° C isothermal; copper-rich alloys.
● T(CoCu) + $CuAl_2$. ∇▼ $\alpha + T$(CoCu) + $CuAl_2$. □■ T(CoCu) + W + $CuAl_2$.

In view of the close limits within which the

$$(\alpha + T(\mathrm{CoCu}) + \mathrm{CuAl_2})/(\mathrm{CuAl_2} + T(\mathrm{CoCu}))$$

boundary is fixed, it is considered justifiable to extrapolate it to obtain an estimate of the composition of T(CoCu) which enters into equilibrium with the α and $CuAl_2$ phases. When extrapolated, this boundary cuts the extrapolated

$$(\alpha + T(\mathrm{CoCu}))/(\alpha + T(\mathrm{CoCu}) + \mathrm{CuAl_2})$$

boundary of figure 7, which is also accurately located, at the composition 13·5 % cobalt, 36·1 % copper, and 50·4 % aluminium. The cobalt figure can hardly lie outside the limits 13·1 to 13·6 %, while the corresponding range for copper is 36·0 to 36·4 %.

This work therefore establishes the presence, in aluminium-rich aluminium-cobalt-copper alloys, of a ternary compound, the composition of which is more fully discussed in § 11.

(c) *The* 500° C *isothermal—aluminium-rich alloys*

Forty-eight aluminium-rich alloys were annealed successively for three periods of 4 weeks, and were examined micrographically at each stage. As in the experiments at 530° C, after the shorter annealing treatments several alloys in the region of the ($\alpha+T$(CoCu)) field contained the four phases α, Co_2Al_9, T(CoCu) and $CuAl_2$, and the effect of further annealing was to reduce the amount of one of these phases. The rate of attainment of equilibrium was, however, lower than at 530° C, and even after a total period of 12 weeks' annealing, alloys 1·32/7·63, 1·46/7·80, 0·76/5·68, 1·00/6·82 (see figure 5), 1·36/7·05, 2·64/7·54, 1·08/6·04, 1·0/11·0, 1·50/10·1, 2·20/9·96, 2·99/9·70 and 3·0/9·0 contained small amounts of a fourth phase. A further annealing treatment of 4 weeks, making a total of 16 weeks, was therefore given to these alloys and to critical three-phase alloys. After this treatment, the seven alloys 1·32/7·63, 1·46/7·80, 0·76/5·68, 1·00/6·82, 1·36/7·05, 2·20/9·96 and 2·99/9·70 still contained small amounts of a fourth phase; the microstructures showed α, T(CoCu) and approximately equal amounts of Co_2Al_9 and $CuAl_2$. It was therefore impossible, without prohibitively long annealing treatments, definitely to decide to which of the neighbouring

$$(\alpha+T(\mathrm{CoCu})+\mathrm{Co_2Al_9}) \quad \text{and} \quad (\alpha+T(\mathrm{CoCu})+\mathrm{CuAl_2})$$

fields these alloys should be assigned. Since the results adequately confirmed the 530° C isothermal, and the presence of the ternary compound T(CoCu), the experiments were discontinued after 16 weeks.

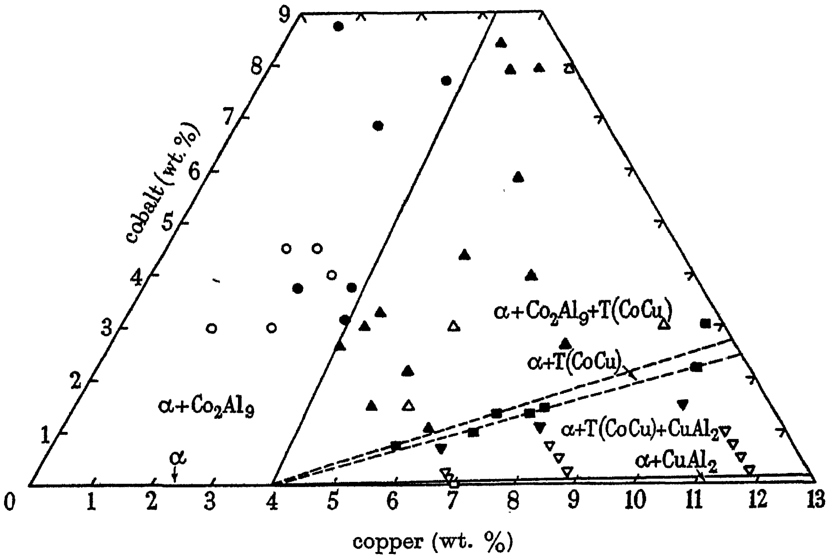


FIGURE 9. Key as for figure 7. The 500° C isothermal; aluminium-rich alloys.

The results obtained are plotted in figure 9. The boundaries of the ($\alpha+T$(CoCu)) field must lie between the points for alloys 1·08/6·04 and 0·7/6·46, and cannot differ significantly from those shown. In view, however, of the persistence of metastable conditions in this region, these boundaries have been drawn as broken lines. The general form of the diagram is similar to that for alloys annealed at 530° C, the only significant difference being caused by the decreased solid solubility of copper in aluminium at 500° C.

8. FURTHER EXPERIMENTS ON THE CONSTITUTION OF THE ALLOYS

The annealing work indicates that, in addition to Co_2Al_9 and $CuAl_2$, a ternary compound T(CoCu) may enter into equilibrium with the primary solid solution, while the occurrence of inclusions of Co_2Al_9 in particles of T(CoCu) suggests that the ternary compound is formed peritectically from Co_2Al_9. This implies that three-phase ($\alpha + T(\mathrm{CoCu}) + Co_2Al_9$) alloys should melt at a higher temperature than three-phase ($\alpha + T(\mathrm{CoCu}) + CuAl_2$) alloys, and experiments were made to test this. Specimens of alloy 1·5/5·5, annealed to equilibrium in the ($\alpha + T(\mathrm{CoCu}) + Co_2Al_9$) field at 530° C, were annealed for 1 hr. at successively increasing temperatures, and examined micrographically after quenching. At 557·0° C, only the three solid phases were visible, but at 560·4° C, the alloy contained the α and Co_2Al_9 phases together with an appreciable amount of liquid. The temperature of the invariant reaction at which T(CoCu) is formed is therefore 558·7 ± 1·7° C. A sample of alloy 0·25/6·75, however, showed the α, $CuAl_2$ and liquid phases after annealing at 553·4° C, indicating that three-phase alloys containing $CuAl_2$ melt at a lower temperature than those containing Co_2Al_9.

Several cooling curves of an alloy containing 1·75 % Co and 14·25 % Cu were taken in order to establish whether the α, T(CoCu) and $CuAl_2$ phases separated from the liquid as a eutectic. A weak arrest was obtained at approximately 667° C, corresponding to the separation of Co_2Al_9. A heavy arrest, corresponding to the secondary separation of α and Co_2Al_9, occurred within the limits 614·5 to 615·2° C, while a final horizontal arrest occurred at 543·9 to 545·1° C. It is therefore established that a ternary eutectic occurs at approximately 544·5° C, 3·5° C below the temperature of the binary aluminium-copper eutectic temperature. No arrest corresponding to the peritectic formation of T(CoCu) was observed either on slow cooling, or on slow heating after annealing the alloy, with stirring, approximately 1° C above the final arrest for 30 min. The peritectic reaction involving the formation of T(CoCu) from Co_2Al_9 is therefore very incomplete under slow cooling conditions.

9. THE SURFACES OF PRIMARY SEPARATION

Several alloys were examined micrographically after slow cooling from the liquid state. The compositions of these alloys are given in figure 10, and the points are distinguished according to the phase which was deposited as primary crystals. In all alloys represented by points in this diagram, a ternary eutectic structure consisting of the α, T(CoCu), and $CuAl_2$ phases was present. The alloys which deposited Co_2Al_9 crystals free from peritectic rims contained secondary α and Co_2Al_9, while those which deposited Co_2Al_9 crystals, upon which T(CoCu) rims subsequently formed by peritectic reaction, contained secondary T(CoCu) and $CuAl_2$. Alloys containing primary crystals of T(CoCu) showed a secondary deposition of T(CoCu) and $CuAl_2$. An alloy containing 0·33 % cobalt and 31·82 % copper (not shown in figure 10) consisted almost entirely of the ternary eutectic; a trace of secondary α and T(CoCu) was visible, so that this composition probably lies slightly to the aluminium-rich side of the eutectic composition.

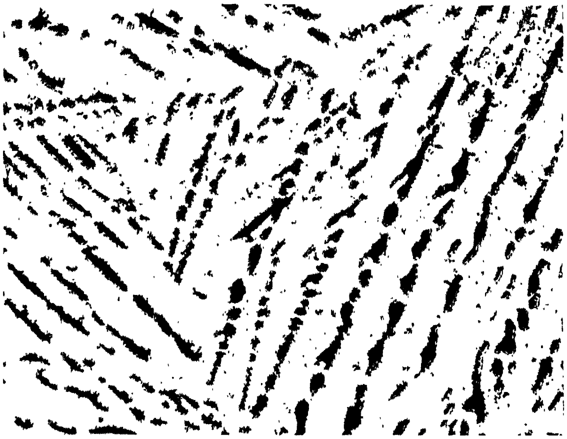


FIGURE 3. Alloy 6·82/2·33, annealed for 8 weeks at 500° C, and etched in sodium hydroxide and sodium carbonate solution. α (background) + Co_2Al_9 (dark). × 500.

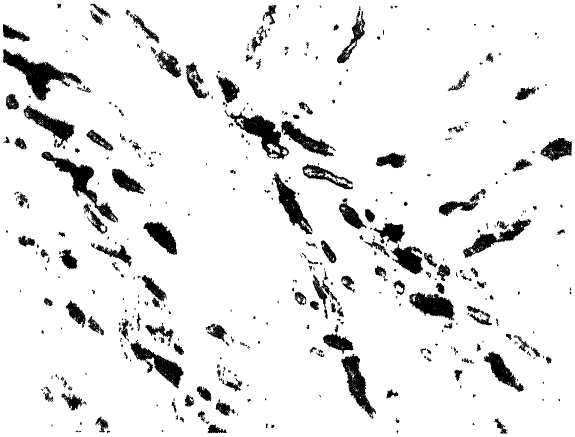


FIGURE 4. Alloy 4·38/5·00, annealed for 8 weeks at 500° C, and etched in sodium hydroxide and sodium carbonate solution. α (background) + Co_2Al_9 (dark) + T(CoCu) (light). × 500.

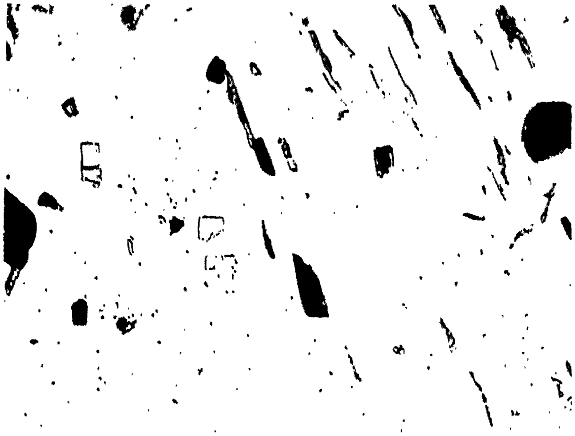


FIGURE 5. Alloy 1·00/6·82, annealed for 12 weeks at 500° C, and etched with 10 % nitric acid after immersion in sodium hydroxide and sodium carbonate solution. α (background) + T(CoCu) (light) + Co_2Al_9 (half-tone) + $CuAl_2$ (black). × 500.

FIGURE 6. Alloy 0·7/6·46, annealed for 12 weeks at 500° C, and etched with 10 % nitric acid after immersion in sodium hydroxide and sodium carbonate solution. α (background) + T(CoCu) (light) + $CuAl_2$ (black). × 500.

The experiments summarized in figure 10 show that the primary Co_2Al_9 field is very extensive, while that for T(CoCu) is very limited. The peritectic relationship between Co_2Al_9 and T(CoCu) is confirmed. The form of the fields of primary separation explains the difficulty of attaining equilibrium by annealing in the solid state. For example, the alloy 1·5/10·5, on casting, would be expected to deposit primary crystals of Co_2Al_9, followed by a secondary deposition of α and Co_2Al_9, and finally to deposit α, T(CoCu) and $CuAl_2$ at the ternary eutectic. Thus, in the 'as-cast' state Co_2Al_9 would be present in relatively large amounts, accompanied by only a little T(CoCu). In the fully annealed condition, however, no Co_2Al_9 is present, but there is much T(CoCu) and some $CuAl_2$. Examination of 'as-cast' specimens of alloy 1·5/10·5 confirmed that much Co_2Al_9, but only a trace of T(CoCu) was present.

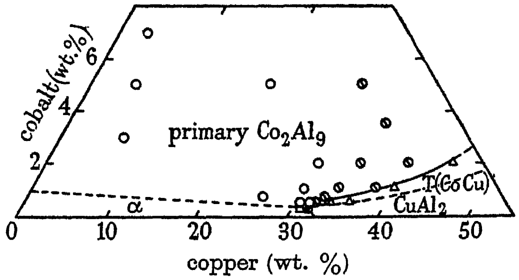


FIGURE 10. Fields of primary separation.

○ Primary Co_2Al_9. ⦸ Co_2Al_9 rimmed by T(CoCu). △ T(CoCu). ● $CuAl_2$. □ α.

For alloys within the range of compositions studied, therefore, Co_2Al_9 will predominate in the 'as-cast' specimens, and the slowness of the attainment of equilibrium on annealing is due essentially to the drastic rearrangement necessary to absorb much or all of this phase, and to form T(CoCu). This illustrates the necessity for adequate annealing in this type of system, and indicates the manner in which too cursory an examination could give erroneous results. If the small amounts of T(CoCu) present in many alloys after short annealing were to escape notice, an alloy containing, at equilibrium, the T(CoCu) and $CuAl_2$ phases could be erroneously assigned to a fictitious ($\alpha + Co_2Al_9 + CuAl_2$) field. Since the recognition of small amounts of a fourth phase in a ternary alloy, and hence of non-equilibrium conditions, is probably easier by the use of the microscope than by the use of X-ray diffraction methods, microscopic methods for this type of system appear preferable.

10. THE ANALYSIS OF EXTRACTED PRIMARY CRYSTALS

In order to investigate the solubility of copper in Co_2Al_9, the primary crystals deposited on slow cooling in alloys 7·0/3·0, 5·0/5·0, 3·0/7·0 and 3·5/35 were extracted electrolytically, and analyzed. To avoid the contamination of the crystals from alloy 3·5/35 with T(CoCu), the alloy was quenched in water from a temperature (570° C) above that corresponding with the peritectic formation of the ternary phase. The analytical results (table 2) confirm that the solubility of copper in Co_2Al_9 is limited, as suggested by the annealing experiments. The solubility is too restricted for the mode of replacement of cobalt by copper to be definitely established.

Several attempts were made to extract primary T(CoCu) crystals from alloy 0·5/33·0. It was found, however, that the large amount of copper present redeposited on the ingot during electrolysis, and led to gross contamination by spongy metal. In addition, the ternary compound crystals themselves were attacked by dilute acids, so that it was possible to obtain only a very few large crystals for examination. They were obtained as long, shiny needles, but in a quantity too small for analysis.

Table 2

composition of alloy		nature of extract	composition of extract	
Co %	Cu %		Cu %	Co %
7·0	3·0	plates, characteristic of Co$_2$Al$_9$	0·20	32·20
5·0	5·0		0·20	33·35
3·0	7·0		0·67	32·00
3·5	35·0		1·44	32·06

Further attempts were made, using an alloy containing 1·75 % cobalt and 14·25 % copper and annealed at 554° C in the $(\alpha + T(\text{CoCu}) + \text{liquid})$ region of the equilibrium model for a sufficient time to allow growth of the T(CoCu) particles to a reasonable size. The extraction was again unsuccessful, and no analysis of the compound T(CoCu) proved possible.

Owing to the very low rate of attainment of equilibrium in this system, and to the existence of successive peritectic reactions in copper-rich alloys containing more than a few per cent of cobalt, it was considered impracticable to prepare pure samples of T(CoCu) by casting and annealing alloys of the requisite composition. It was thus not possible to supplement the information with regard to the composition of T(CoCu) obtained from the annealing experiments described in § 8.

11. Discussion

The work described in this paper establishes that, as in the systems aluminium-iron-copper and aluminium-nickel-copper, a ternary compound enters into equilibrium with the primary solid solution in the aluminium-cobalt-copper alloys.

Although the compounds which occur in the iron-bearing and nickel-bearing alloys have been described respectively as FeCu$_2$Al$_7$ (Bradley & Lipson 1938) and NiCu$_3$Al$_6$ (Bradley & Goldschmidt 1939), it is clear from the published constitutiona diagrams that they exist over appreciable ranges of composition and hence of electron : atom ratio. The data with regard to the ranges of homogeneity are summarized in table 3; in this table, electron : atom ratios have been calculated on the basis of the theory outlined above, that is, transitional metal atoms have been regarded as absorbing sufficient electrons from the structure as a whole to fill the vacancies in their atomic orbitals. The characteristic electron : atom ratios are only slightly modified if the effective valencies of nickel and iron are taken as zero and -2 respectively. Figure 11 shows how the range of homogeneity in each case is distributed on either side of the characteristic electron : atom ratio of approximately 2·04. According to table 3, these phases may exist over the range of electron : atom ratios from 1·96 to 2·14.

In figure 12, which represents the aluminium-cobalt-copper system, lines have been drawn at electron : atom ratios of 1·96 and 2·14; the point representing the composition of T(CoCu) which enters into equilibrium with the α and $CuAl_2$ phases is also plotted, and it is seen that this falls only slightly outside the critical range of electron : atom ratios for the phases $FeCu_2Al_7$ and $NiCu_3Al_6$. It is therefore very probable that T(CoCu) is to be regarded as analogous to these phases, and based upon the same electron : atom ratio.

TABLE 3

	compound $FeCu_2Al_7$			
	Fe	Cu	Al	electron: atom ratio
composition in equilibrium with α and $CuAl_2$ (at. %)	8·75	18·40	72·85	2·14
composition in equilibrium with α and $FeAl_3$ (at. %)	11·20	15·00	73·80	2·07
compositions at apices of other three-phase triangles	11·90	18·00	70·10	1·97
	11·00	20·75	68·25	1·96
	10·00	21·75	68·25	2·00
	9·00	22·25	68·75	2·04

	compound $NiCu_3Al_6$			
	Ni	Cu	Al	electron: atom ratio
composition in equilibrium with α and $CuAl_2$ (at. %)	8·70	30·87	60·43	2·07
composition in equilibrium with α and $NiAl_3$ (at. %)	12·39	25·65	61·96	2·04
compositions at apices of other three-phase triangles	4·13	37·39	58·48	2·10
	6·96	36·96	56·08	2·01
	8·70	36·09	55·21	1·96
	11·52	32·39	56·09	1·94

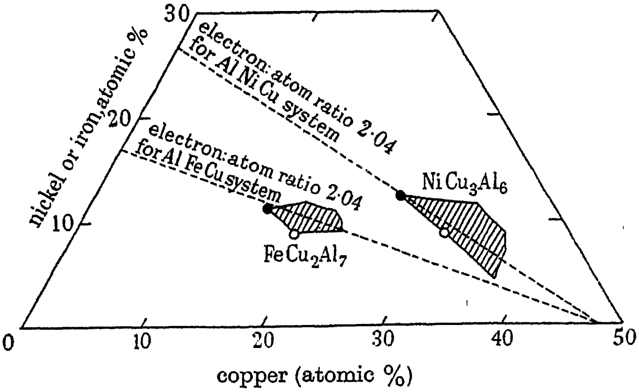


FIGURE 11. Homogeneity ranges of $FeCu_2Al_7$ and $NiCu_3Al_6$.

○ Composition in equilibrium with α and $CuAl_2$. ● Composition in equilibrium with α and $NiAl_3$ or $FeAl_3$.

If it is assumed that $NiCu_3Al_6$ represents the arrangement upon which the aluminium-copper-nickel phase is based, and that the total number of atoms in this arrangement remains constant when other transitional metal atoms are substituted for nickel, the characteristic formula for the aluminium-iron-copper phase can readily be understood. The substitution of iron for nickel in the structure increases the number of vacancies in the transitional metal atoms by approximately two. To maintain the same electron : atom ratio, the copper and aluminium atoms must together supply two more electrons than in the aluminium-nickel-copper compound. Since copper and aluminium are respectively mono- and trivalent, this can only be accomplished by adding one aluminium atom and subtracting one copper atom if the total number of atoms present is to remain unaltered. This leads to the characteristic formula $FeCu_2Al_7$. Similar reasoning for the aluminium-cobalt-copper ternary phase leads to the formula $Co_2Cu_5Al_{13}$, since cobalt has approximately one more vacancy per atom than nickel.

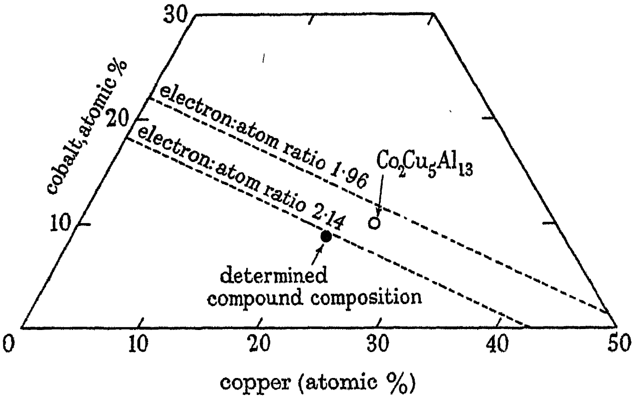


FIGURE 12. The ternary compound in the aluminium-cobalt-copper alloys.

The composition of T(CoCu) which enters into equilibrium with the α and $CuAl_2$ phases corresponds almost exactly with the formula $Co_2Cu_5Al_{16}$; it will be appreciated that the atomic ratio of copper to cobalt is accurately established by the work summarized in figure 7, and is in exact agreement with the theoretical expectations discussed above. The atomic ratio of aluminium to cobalt is not completely established by the present work, since only the composition of T(CoCu) in equilibrium with the α and $CuAl_2$ phases has been determined. Since, however, the compound probably has a range of homogeneity corresponding with those for the $FeCu_2Al_7$ and $NiCu_3Al_6$ phases, it is also probable that it can exist with a smaller number of aluminium atoms than corresponds to the experimentally determined formula. The composition $Co_2Cu_5Al_{13}$, which would be completely analogous to those for the other two compounds, falls well within the critical range of electron : atom ratios, as shown in figure 12.

It is of interest to note that the ratio of copper atoms to transitional metal atoms rises in a regular manner in the three compounds as the number of vacancies in the atomic orbitals of the transitional metal decreases, while at the same time, the ratio of aluminium atoms to transitional metal atoms decreases. The latter tendency is entirely analogous to the manner in which the ratio of aluminium atoms to transitional metal atoms decreases in the series of binary compounds $CrAl_7$, $MnAl_6$, Co_2Al_9 and $NiAl_3$.

The results obtained from the examination of the aluminium-cobalt-copper alloys therefore demonstrate that a ternary compound enters into equilibrium with the primary solid solution based on aluminium, and that this ternary compound is to be considered as analogous to those which exist in the ternary alloys of aluminium and copper with iron and nickel respectively. Support is given to the view that, in aluminium-rich alloys, transitional metals of the first long period absorb electrons from the structure to an extent governed by the number of vacancies per atom in their atomic orbitals (or in their 3*d* electron shells), and that ternary compound phases in aluminium-rich alloys are markedly influenced by factors of an electronic origin.

This research was carried out under the general supervision of Professor D. Hanson, D.Sc., to whom the authors' thanks are due for his interest and support. The authors must also acknowledge the valuable assistance of Mr G. Welsh in the experimental work, and the analytical work of Mr P. Wingrove. Grateful acknowledgement is due to the Department of Scientific and Industrial Research, The Royal Society, the Chemical Society and Imperial Chemical Industries, Ltd. for general financial assistance.

REFERENCES

Bradley, A. J. & Goldschmidt, H. J. 1939 *J. Inst. Metals*, **65**, 403.
Bradley, A. J. & Lipson, H. 1938 *Proc. Roy. Soc.* A, **167**, 421.
Fink, W. L. & Freche, H. R. 1932 *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **99**, 141.
Pauling, L. 1938 *Phys. Rev.* (ii), **54**, 899.
Raynor, G. V. 1944*a* *J. Inst. Metals*, **70**, 531.
Raynor, G. V. 1944*b* *Inst. Metals Annotated Equilibrium Diagrams Series*, no. 4.
Raynor, G. V. 1945 *Phil. Mag.* (vii), **36**, 770.
Raynor, G. V. & Little, K. 1945 *J. Inst. Metals*, **71**, 493.
Raynor, G. V. & Pfeil, P. C. L. 1946–7*a* *J. Inst. Metals*, **73**, 387.
Raynor, G. V. & Pfeil, P. C. L. 1946–7*b* *J. Inst. Metals*, **73**, 609.
Raynor, G. V. & Wakeman, D. W. 1947 *Proc. Roy. Soc.* A, **190**, 82.
Raynor, G. V. & Wakeman, D. W. 1948 *Phil. Mag.* (vii), **39**, 245.
Raynor, G. V. & Waldron, M. B. 1948 *Proc. Roy. Soc.* A, **194**, 362.

The refinement of atomic parameters by the technique known in X-ray crystallography as 'the method of steepest descents'

By A. D. Booth, *Birkbeck College, London*

(*Communicated by J. D. Bernal, F.R.S.—Received* 4 *June* 1948—
Revised 8 *October* 1948)

This paper is the first of a proposed set of three in which applications of the methods of 'steepest descents' are made to the problems of crystal structure analysis. The proposed programme is in three parts:

(1) Working details of the method.
(2) A discussion of the accuracy and convergence of the method.
(3) Examples of the use of the method in actual structure determination.

The present paper forms the first of this set and is complete in itself.

A discussion of the normal methods of structure analysis results in a formulation of the problem as one of minimization, and leads naturally to the idea of optimum procedures for this purpose.

Three techniques are derived from a unified presentation: (*a*) the 'relaxation' method, (*b*) the 'least squares' refinement, (*c*) the method of 'steepest descents'. These are examined in detail and their virtues and defects noted.

A complete discussion of the application of the steepest descent procedure to various types of structure analysis is given and techniques are derived for the cases of: refinement of a set of parameters already reasonably good; refinement of a structure where general configuration but *not* orientation is known; and refinement of structures given by electron densities on a lattice.

Several variants of the process are given, in one of which only those planes of *zero* observed intensity being used; this gives a process which is independent of phase angle values and should be useful in the case of structures for which no chemical data is available.

1. Introduction

Since their introduction into the field of practical structure analysis by Bragg (1929), multi-dimensional Fourier syntheses have held first place as a means of deriving accurate atomic parameters from X-ray data. From the early and simple structures first examined, the method has been developed into a powerful analytical tool which has proved capable of obtaining results (as in penicillin) where the methods of organic chemistry have been unsuccessful.

With the increasing complexity of the molecules examined, however, it has become more and more evident that despite technical improvements in the computing techniques (e.g. Booth 1945, 1946*a*, *b*), the time is rapidly approaching when the method will prove inadequate.

The main reason for this decline is that an increase in the number of parameters (i.e. atoms) in a structure produces a disproportionately greater increase in the labour of finding a trial structure to commence the process of Fourier refinement.

2. Statement of the problem

It is worth while setting out the precise requirements of a structure determination in mathematical form.

A set of observations on the X-ray diffraction pattern of a given crystal results in the experimental determination of quantities $F_o(h,k,l)$ $(h,k,l,$ integers).

These constitute the moduli of a set of coefficients $F_c(h,k,l)$ defined by the relation

$$F_c(h,k,l) = \sum_r f_r \exp\left[2\pi i(hx_{r/a} + ky_{r/b} + lz_{r/c})\right], \tag{2·1}$$

where f_r = atomic scattering factor of rth atom, and (x_r, y_r, z_r) = co-ordinates of rth atom in unit cell.

The problem of structure analysis is, given the F_o, determine the (x_r, y_r, z_r). Mathematically equivalent, although subject to modification later, is the statement—determine (x_r, y_r, z_r) such that

$$R_n = \sum_{hkl} G(g(F_o) - g(F_c)) \text{ is a minimum,}$$

the G and g functions being arbitrary except that $G \geqslant 0$. For example,

$$R = \Sigma\,\big|\,|F_o| - |F_c|\,\big| \quad R_1 = \Sigma(|F_o| - |F_c|)^2 \quad R_2 = \Sigma(F_o^2 - F_c^2)^2$$

may be minimized. The term minimum is used in preference to zero since, because of the experimental errors in the $|F_o|$, this is never attained.

The above statement requires modification because a set of (x_r, y_r, z_r) may satisfy the requirement of minimizing R and still be untenable on chemical grounds. It is thus necessary to introduce further conditions of the type

$$(x_r - x_s)^2 + (y_r - y_s)^2 + (z_r - z_s)^2 + 2(x_r - x_s)(y_r - y_s)\cos\gamma + \text{etc.} = l_{rs}^2, \tag{2·2}$$

where l_{rs} is the distance between the atoms r and s and is assumed to be known on chemical grounds.

Now although the conditions (2·2) may be needed to establish a unique structure of the correct chemical configuration, they must be removed for the final refinement, since it is often the object of this to determine the bonds—or more correctly the configuration constants—l_{rs}.

It is thus evident that at least three methods of approach are needed:

(1) To apply when the atomic positions are approximately known.

(2) To apply when the configuration of the molecule is known in detail but not its position in the unit cell.

(3) To apply when the general molecular shape is known but none of its details.

In the present paper it is proposed to show how the method of 'steepest descents' offers a solution to each of these problems, and to develop the detailed procedures from the basic assumptions of the method. Besides this it will be shown how the Southwell 'relaxation' technique and the 'least squares' refinement form variations on the general theme. It is intended, in a further paper, to examine the question of relative accuracy and of convergence.

Prior to giving an account of the methods, however, it is proposed to explain the method of 'steepest descents' which forms the basis of the treatment.

3. THE METHOD OF STEEPEST DESCENTS

For the purpose of illustration, consider a set of quantities $F_c(h,k,l)$ defined by the equations

$$F_c(h,k,l) = \psi(x,y,h,k,l), \tag{3·1}$$

and suppose that, given another set of quantities $F_o(h, k, l)$, it is required to determine (x, y) so that the quantity

$$R_2 = \sum_{h,k,l} (F_o^2 - F_c^2)^2 \tag{3·2}$$

is a minimum.

Now using (3·1) and (3·2) it follows that

$$R_2 = \phi(x, y), \tag{3·3}$$

where ϕ is a function determinable when the form of ψ is known.

In the two-dimensional case given above it is possible to draw a contour map of R_2 as a function of (x, y), and it will appear as a series of regions of the type given in figure 1. If any point $P_0(x, y)$ be chosen at random the value of R_2 will be, in general, different from its minimum value. It is required to define a procedure which will enable the point at which R_2 is a minimum to be located.

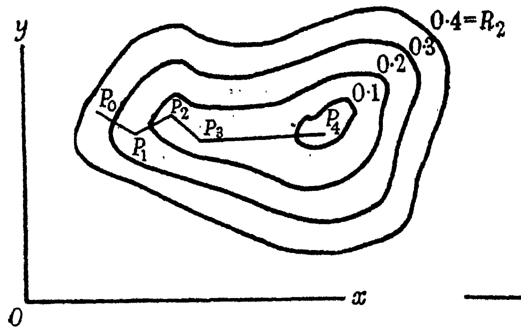


FIGURE 1

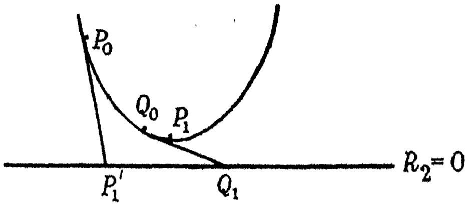


FIGURE 2

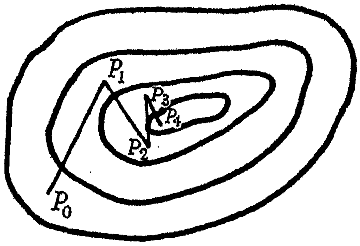


FIGURE 3

A method of achieving this is as follows. Starting at P_0 determine the direction of greatest slope of the function $\phi(x, y)$ and proceed down it until R_2 ceases to decrease; this gives the point P_1. The procedure is then repeated at P_1 to give P_2 and again at P_2 to give P_3. The process obviously converges to the minimum of R_2.

An alternative method, which gives good results in the early stages of the process, is the following. From P_0 the direction of steepest descent is followed until it reaches the plane $R_2 = 0$. The co-ordinates of the point P_1' so defined are taken as the origin for the next refinement.

This is shown in section in figure 2. The defect of the method lies in the fact that if P_0 is near to the minimum P_1 (as at Q_0), the intersection point Q_1 may overshoot to a very large extent so that the value of R_2 at Q_1 is in fact greater than that at Q_0.

The seriousness of this situation becomes apparent when a region of the type shown in figure 3 is considered. In this case the overshooting greatly increases the number of refinements which have to be made. Since, however, the latter method is much the simplest to apply, it has been found useful in the initial stages of a refinement and will be described.

4. ANALYTICAL TREATMENT

To avoid duplication of cumbersome equations the notation

$$\left.\begin{aligned} x_{rj} &= x_r \quad (j=1),\\ &= y_r \quad (j=2),\\ &= z_r \quad (j=3) \end{aligned}\right\} \tag{4·1}$$

will be used.

Now if the relation $$R_2 = \phi(x_{1,1} \dots x_{r,j} \dots x_{N,3}) \tag{4·2}$$

obtains, the generalized direction cosines of the normal to the surface $R_2 = \text{const.}$, in the $3N$-dimensional space defined by the co-ordinates $(x_{r,j})$, are given by

$$\cos(n, x_{rj}) = \frac{\partial\phi}{\partial x_{rj}}\bigg/\frac{\partial\phi}{\partial n} \quad (r = 1 \dots N;\ j = 1, 2, 3), \tag{4·3}$$

where $$\frac{\partial\phi}{\partial n} = |\operatorname{grad}\phi| = \left\{\sum_{r,j}\left(\frac{\partial\phi}{\partial x_{rj}}\right)^2\right\}^{\frac{1}{2}}. \tag{4·4}$$

If the point $({}_0x_{rj})$ $(r = 1 \dots N;\ j = 1, 2, 3)$ is taken as the initial point then three expansions of generalized Taylor type are possible:

$$R_2' = {}_0R_2 + \epsilon_{rj}\frac{\partial\phi}{\partial x_{rj}} + \frac{1}{2!}\epsilon_{rj}^2\frac{\partial^2\phi}{\partial x_{rj}^2} + \text{etc.}, \tag{4·5}$$

$$R_2' = {}_0R_2 + \sum_{rj}\epsilon_{rj}\frac{\partial\phi}{\partial x_{rj}} + \frac{1}{2!}\sum_{r,j;\,r',j'}\epsilon_{r,j}\,\epsilon_{r'j'}\frac{\partial^2\phi}{\partial x_{rj}\,\partial x_{r'j'}} + \text{etc.}, \tag{4·6}$$

$$R_2' = {}_0R_2 + \epsilon_n\frac{\partial\phi}{\partial n} + \frac{1}{2!}\epsilon_n^2\frac{\partial^2\phi}{\partial n^2} + \text{etc.} \tag{4·7}$$

Here R_2' is the value of R_2 at the point $(x_{rj} + \epsilon_{rj})$ $(r = 1 \dots N;\ j = 1, 2, 3)$ and the ϵ_{rj} and ϵ_n are considered to be small.

The three expansions lead to rather different results; (4·5) gives the Southwell 'relaxation' technique, (4·6) the 'least squares' refinement, and (4·7) the 'steepest descent' methods.

5. THE RELAXATION METHOD

In equation (4·5) it is assumed that only one parameter (x_{rj}) is being varied at a time. If the variation of R_2 with ϵ_{rj} is considered its form will be similar to that shown in the graph of figure 2. Supposing the initial point to be not too near the minimum, the point P_1' will correspond to a smaller value of R_2 than P_0, and the value of ϵ_{rj} required to reach this is obtained by neglecting squares and higher of ϵ_{rj} in (4·5), giving

$$\epsilon_{rj} = -\frac{{}_0R_2}{\partial\phi/\partial x_{rj}}. \tag{5·1}$$

When P_0 is already near to its minimum, however, the derivative in (5·1) is very small and ϵ_{rj} may thus become so large that overshooting to a value of R_2 larger than initially will occur. In this case a second-order term must be included and the *minimum* value of R_2 determined. The value of ϵ_{rj} required for this is obtained from the minimization condition

$$\frac{\partial R_2}{\partial \epsilon_{rj}} = 0,$$

or, substituting from (4·5),

$$\frac{\partial \phi}{\partial x_{rj}} + \epsilon_{rj}\frac{\partial^2 \phi}{\partial x_{rj}^2} = 0,$$

which gives

$$\epsilon_{rj} = -\frac{\partial \phi}{\partial x_{rj}} \Big/ \frac{\partial^2 \phi}{\partial x_{rj}^2}. \tag{5·2}$$

Although the above equations (5·1) and (5·2) are very simple and the resulting values of ϵ_{rj} easy to compute, the method suffers from several severe disadvantages and limitations of which the chief may be mentioned.

In the first place since the method always descends parallel to one of the axes, this direction tends, with increasing number of parameters, to be *perpendicular* to the direction of *steepest* descent. (Actually $\cos^{-1}(1/\sqrt{n})$, where n = number of parameters.) This may lead, even in the simplest cases, to very slow convergence; consider, for example, the two-dimensional case represented in figure 4. Here it is evident that a very large number of steps would be required in order to reach the minimum by any axial descent method, whereas one steepest descent from P_1 would lead to the major axis of the elliptical contours, and a second along this axis to the minimum.

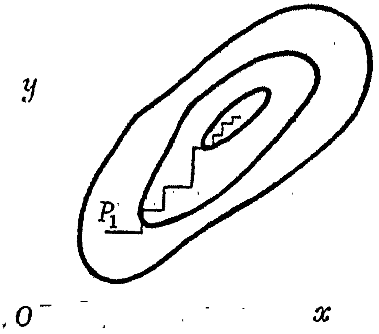


FIGURE 4

6. THE LEAST SQUARES REFINEMENT

Suppose that it is required to minimize the value of R_2' as given by (4·6). The requirements are simply

$$\frac{\partial R_2'}{\partial \epsilon_{rj}} = 0 \quad (r = 1 \dots N;\ j = 1, 2, 3). \tag{6·1}$$

The application of these relations to (4·6) gives the set of $3N$ simultaneous equations

$$\sum_{rj \geqslant r'j'} \epsilon_{r'j'} \frac{\partial^2 \phi}{\partial x_{rj}\,\partial x_{r'j'}} + \frac{\partial \phi}{\partial x_{rj}} = 0 \quad (r = 1 \dots N;\ j = 1, 2, 3), \tag{6·2}$$

which would give on solution the values of ϵ_{rj} required to minimize R_2. These are precisely the normal equations of the least-squares refinement and have been used by Hughes (1941) and by the author in structure determinations.

The method suffers the disadvantage that when N is at all large the labour of solution becomes prohibitive, and in any case, the neglect of higher order terms makes the method only approximate, so that several refinements have to be made unless the initial point (x_{rj}) is very close to the minimum.

7. The method of steepest descents

It is again necessary to describe the two variants envisaged in § 3 (figure 2). The first corresponds to neglecting ϵ_n^2 in (4·7). The value of ϵ_n required to make $R_2 = 0$ is then given by

$$\begin{aligned}\epsilon_n &= -\frac{{}_0R_2}{\partial\phi/\partial n} \\ &= -{}_0R_2\bigg/\left\{\sum_{rj}\left(\frac{\partial\phi}{\partial x_{rj}}\right)^2\right\}^{\frac{1}{2}}. \qquad (7\cdot1)\end{aligned}$$

The individual ϵ_{rj} are easily obtained from the equations

$$\begin{aligned}\epsilon_{rj} &= \epsilon_n \cos(n, x_{rj}), \qquad (7\cdot2) \\ &= -{}_0R_2\left(\frac{\partial\phi}{\partial x_{rj}}\right)\bigg/\sum_{rj}\left(\frac{\partial\phi}{\partial x_{rj}}\right)^2, \qquad (7\cdot3)\end{aligned}$$

by (4·3) and (4·4).

The more sophisticated method proceeds to the minimum of the parabola defined by $R_2 = f(\epsilon_n)$ in (4·7) and gives

$$\epsilon_n = -\left(\frac{\partial\phi}{\partial n}\right)\bigg/\frac{\partial^2\phi}{\partial n^2},$$

or, using the relation

$$\begin{aligned}\frac{\partial^2\phi}{\partial n^2} &= \frac{\partial}{\partial n}\left(\frac{\partial\phi}{\partial n}\right) = \sum_{rj}\frac{\partial}{\partial x_{rj}}\left(\frac{\partial\phi}{\partial n}\right)\frac{\partial x_{rj}}{\partial n} \\ &= \sum_{rj,\,r'j'}\frac{\partial^2\phi}{\partial x_{rj}\,\partial x_{r'j'}}\frac{\partial\phi}{\partial x_{rj}}\frac{\partial\phi}{\partial x_{r'j'}}\bigg/\sum_{rj}\left(\frac{\partial\phi}{\partial x_{rj}}\right)^2,\end{aligned}$$

whence

$$\epsilon_n = -\left[\sum_{rj}\left(\frac{\partial\phi}{\partial x_{rj}}\right)^2\right]^{\frac{3}{2}}\bigg/\sum_{rj,\,r'j'}\frac{\partial^2\phi}{\partial x_{rj}\,\partial x_{r'j'}}\frac{\partial\phi}{\partial x_{rj}}\frac{\partial\phi}{\partial x_{r'j'}}, \qquad (7\cdot4)$$

and from (7·2)

$$\epsilon_{rj} = -\left[\sum_{rj}\left(\frac{\partial\phi}{\partial x_{rj}}\right)^2\right]\frac{\partial\phi}{\partial x_{rj}}\bigg/\sum_{rj,\,r'j'}\frac{\partial^2\phi}{\partial x_{rj}\,\partial x_{r'j'}}\frac{\partial\phi}{\partial x_{rj}}\frac{\partial\phi}{\partial x_{r'j'}}. \qquad (7\cdot5)$$

Equations (7·3) and (7·5) contain the solution of the steepest descent problem in both the simplified and complete forms.

8. Geometrical considerations

It is interesting to consider the behaviour of the approximate method of (7·3) in some detail.

The parabolic relation followed by R_2' considered as a function of ϵ_n is shown in figure 5. The simplified method is equivalent to following the tangent at P to this parabola until it meets the ϵ_n axis in A. Now it is easily seen that

$$OA = -{}_0R_2\bigg/\left(\frac{\partial\phi}{\partial n}\right), \qquad (8\cdot1)$$

and that the abscissa, M, of the true minimum is

$$\epsilon_n(\text{min.}) = -\left(\frac{\partial\phi}{\partial n}\right)\Big/\frac{\partial^2\phi}{\partial n^2}. \tag{8·2}$$

The ordinate of this minimum is

$$R_2(\text{min.}) = {}_0R_2 - \frac{1}{2}\left(\frac{\partial\phi}{\partial n}\right)^2\Big/\frac{\partial^2\phi}{\partial n^2}, \tag{8·3}$$

so that if $R_2(\text{min.}) = 0$,

$$\left(\frac{\partial\phi}{\partial n}\right)\Big/\frac{\partial^2\phi}{\partial n^2} = 2\,{}_0R_2\Big/\frac{\partial\phi}{\partial n}, \tag{8·4}$$

or, substituting in (8·2),

$$\epsilon_n(\text{min.}) = -2\,{}_0R_2\Big/\frac{\partial\phi}{\partial n}. \tag{8·5}$$

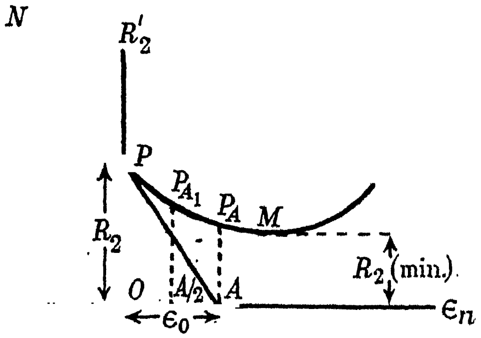


FIGURE 5

This means that the value of ϵ_n, given by the naïve method and typified by OA in (8·1), is half the required value, and that, in these circumstances, the value given by this procedure should be doubled. Although, in practice, experimental errors will always make R_2(min.) different from zero, this property can be used in the early stages of the refinement.

Another useful property of the tangent method becomes apparent when the point P coincides with an inflexion in the (R_2, ϵ_n) curve. In these circumstances the value of $(\partial^2\phi/\partial n^2)$ is zero, so that the descent-to-a-minimum method cannot be used; the tangent descent method, however, still gives a result. Again, in the case where P is nearer to a maximum (N in figure 5) than to a minimum, the more general method will converge to the maximum value, whereas the simpler method will still converge to the minimum.

These considerations suggest that the proper procedure is to apply the simple method until it ceases to produce smaller values of R_2, and then to continue the refinement by means of the more sophisticated technique.

9. APPLICATION TO REFINEMENT

When the positions of the atoms in a structure are known to a fair accuracy the steepest descent method can be applied, as indicated, to refine the parameters to the optimum set required to minimize R_2.

It is worth remarking that, although the function R_2 was chosen in accord with (3·2), the method will work almost equally well with the function

$$R = \Sigma\,\big|\,|F_o| - |F_c|\,\big|. \tag{9·1}$$

The only disadvantage is that the derivatives of R may be discontinuous in the region of expansion in a series ϵ_n, and this could lead to error. It is found in practice, however, that this does not lead to trouble, and since, for centro-symmetric space groups (9·1) is much more readily computed than (3·2), it should be borne in mind for possible application. The sources of difficulty will be indicated later in this section.

In general,

$$|F_c(h,k,l)|^2 = [\sum_r f_r \cos\theta_r]^2 + [\sum_r f_r \sin\theta_r]^2, \tag{9·2}$$

where

$$\theta_r = 2\pi\left(\frac{h}{a}x_r + \frac{k}{b}y_r + \frac{l}{c}z_r\right), \tag{9·3}$$

or, alternatively,

$$|F_c(h,k,l)| = \sum_r f_r \cos(\theta_r - \alpha), \tag{9·4}$$

where

$$\tan\alpha = \sum_r f_r \sin\theta_r / \sum_r f_r \cos\theta_r. \tag{9·5}$$

LEMMA. In the following argument the quantities $\partial F_c/\partial x_{sj}$ and $\partial\alpha/\partial x_{sj}$ will be required, and it is proposed to detail them here. The notation

$$\left.\begin{aligned} h_j &= h, & a_j &= a, & x_{sj} &= x_s, & \text{if } j &= 1,\\ &= k, & &= b, & &= y_s, & \text{if } j &= 2,\\ &= l, & &= c, & &= z_s, & \text{if } j &= 3, \end{aligned}\right\} \tag{9·6}$$

will be used. Differentiating (9·2) and using (9·4) and (9·5),

$$\frac{\partial|F_c|}{\partial x_{sj}} = -2\pi\frac{h_j}{a_j}f_s \sin(\theta_s - \alpha). \tag{L. 1}$$

Again, from (9·5)

$$\frac{\partial\alpha}{\partial x_{sj}} = 2\pi\frac{h_j}{a_j}\frac{f_s}{|F_c|}\cos(\theta_s - \alpha). \tag{L. 2}$$

Whence, from (L. 1) and (L. 2),

$$\frac{\partial^2|F_c|}{\partial x_{sj}\,\partial x_{s'j'}} = -4\pi^2\frac{h_j h_{j'}}{a_j a_{j'}}f_s\left[\delta_{ss'} - \frac{f_{s'}}{F_c}\cos(\theta_{s'} - \alpha)\right]\cos(\theta_s - \alpha), \tag{L. 3}$$

where

$$\begin{aligned} \delta_{ss'} &= 1 \quad (s = s'),\\ &= 0 \quad (s \neq s'). \end{aligned}$$

Now (9·1) may be written

$$R = \sum_{h,k,l}\big|\,|F_o| - \sum_r f_r \cos(\theta_r - \alpha)\,\big|, \tag{9·7}$$

and, using (L. 1),

$$\frac{\partial R}{\partial x_{sj}} = \frac{2\pi}{a_j}\sum_{h,k,l}\frac{|F_o| - |F_c|}{\big|\,|F_o| - |F_c|\,\big|}h_j f_s \sin(\theta_s - \alpha). \tag{9·8}^*$$

Similarly, differentiating again using (L. 2) and (L. 3),

$$\frac{\partial^2 R}{\partial x_{sj}\,\partial x_{s'j'}} = \frac{4\pi^2}{a_j a_{j'}}\sum_{h,k,l}\frac{|F_o| - |F_c|}{\big|\,|F_o| - |F_c|\,\big|}h_j h_{j'} f_s\left[\delta_{ss'} - \frac{f_{s'}}{F_c}\cos(\theta_{s'} - \alpha)\right]\cos(\theta_s - \alpha). \tag{9·9}$$

---

* Since the function $|F|$ is discontinuous the derivatives may not exist everywhere. The equality sign is therefore only conditionally valid.

In practice it is found that the refinement takes place very well using the formulae (9·8) and (9·9), and, in view of their relative simplicity, it is a good plan to minimize R rather than R_2 in the early stages of a refinement.

If R_2 is used the following treatment is appropriate:

$$R_2 = \sum_{h,k,l} (|F_o|^2 - |F_c|^2)^2,$$

whence
$$\frac{\partial R_2}{\partial x_{sj}} = \frac{8\pi}{a_j} \sum_{h,k,l} (|F_o|^2 - |F_c|^2) F_c h_j f_s \sin(\theta_s - \alpha), \tag{9·10}$$

and
$$\frac{\partial^2 R_2}{\partial x_{sj}\,\partial x_{s'j'}} = -\frac{16\pi^2}{a_j a_{j'}} \sum_{h,k,l} \Big\{ (|F_o|^2 - 3|F_c|^2) h_j h_{j'} f_s f_{s'} \sin(\theta_s - \alpha) \sin(\theta_{s'} - \alpha)$$
$$- (|F_o|^2 - |F_c|^2) |F_o| h_j h_{j'} f_s \left[\delta_{ss'} - \frac{f_{s'}}{F_c} \cos(\theta_{s'} - \alpha) \right] \cos(\theta_s - \alpha) \Big\}. \tag{9·11}$$

Equations (9·8), (9·9), (9·10) and (9·11) are sufficient, in conjunction with (7·3) or (7·5), to determine the directions and magnitudes of steepest descent in the case of refinement of parameters already known approximately. It should be noted that

$$\partial\phi/\partial x_{sj} \equiv \partial R_2/\partial x_{sj}, \quad \partial^2\phi/\partial x_{sj}\,\partial x_{s'j'} \equiv \partial^2 R_2/\partial x_{sj}\,\partial x_{s'j'}.$$

It will be seen from the above that, not only are the analytic expressions for the second derivatives rather complicated, but also that a large number $[\frac{1}{2}n(n-1)]$ have to be calculated. This suggests the extreme desirability of devising a procedure which does not involve them. The simple tangent descent method of (7·3) cannot be said to fill this want directly, but it can be modified to give an excellent practical computing technique.

Suppose that the value of R (or R_2) be calculated at A (figure 5) and also at the point P_{A_1} having as co-ordinates $\epsilon_n \equiv \frac{1}{2}\epsilon_{rj}$ ($r = 1 \dots N$; $j = 1, 2, 3$), it is then a simple matter to interpolate or extrapolate for the true minimum. Thus let

$$R = R_0 \quad \text{at origin,}$$
$$R = R_1 \quad \text{at } \epsilon_n \ (\equiv \epsilon_{rj}),$$
$$R = R_{\frac{1}{2}} \quad \text{at } \tfrac{1}{2}\epsilon_n \ (\equiv \tfrac{1}{2}\epsilon_{rj}).$$

Then it is easily shown, by elementary co-ordinate geometry, that if the curve of descent is assumed to be parabolic the co-ordinates of the minimum are given by

$$\epsilon_{\text{min.}} = (3R_0 + R_1 - 4R_{\frac{1}{2}})/4(R_0 + R_1 - 2R_{\frac{1}{2}}) \times \epsilon_n,$$

so that
$$\epsilon_{rj\text{min.}} = (3R_0 + R_1 - 4R_{\frac{1}{2}})/4(R_0 + R_1 - 2R_{\frac{1}{2}}). \tag{9·12}$$

Care must be taken that ascent to a maximum does not occur; this is, however, easily prevented and the following routine has been found 'safe' in practice:

(1) Compute ${}_0R$ (or ${}_0R_1$ or ${}_0R_2$).

(2) Compute all $\partial R/\partial x_{rj}$.

(3) Find all ϵ_{rj} using (7·3).

(4) Recalculate R.

(5*a*) If $R < {}_0R$, accept the ϵ_{rj} and repeat from (1) using the new co-ordinates $x_{rj} + \epsilon_{rj}$.

(5*b*) If $R > {}_0R$ but the ϵ_{rj} are reasonably small ($< 0{\cdot}1$ Å), calculate the value of R for corrections $\frac{1}{2}\epsilon_{rj}$.

(6*b*) Interpolate for correct ϵ_{rj} using (9·12).

(7*b*) Recalculate R using ϵ_{rj}(min.) and repeat from (1) using this set of $x_{rj} + \epsilon_{rj}$(min.).

(5*c*) If $R > {}_0R$ but the ϵ_{rj} are very large ($> 0{\cdot}1$ Å), scale the ϵ_{rj} down so that the *largest* is of the order of 0·1 Å.

(6*c*) Recalculate the value of R for these ϵ_{rj}.

(7*c*) Calculate R for $\frac{1}{2}\epsilon_{rj}$.

(8*c*) Interpolate for ϵ_{rj}(min.).

(9*c*) Recalculate R using ϵ_{rj}(min.) and repeat from (1).

10. DETERMINATION OF MOLECULAR POSITIONS

Although a set of atoms, placed at random in the unit cell and then varied in position so as to minimize R or R_2, does not, in general, give a unique result or even a structure possible on chemical grounds, it is easy, by assuming the configuration of the molecule, to find the position of this configuration which minimizes R or R_2. In this section, the technique of this process will be developed.

It may be convenient to use the $\frac{1}{2}N(N-1)$ configurational equations (2·2) to establish the rigidity of a molecular arrangement, but sometimes a better method is the following:

(1) Evaluate co-ordinates of atomic positions in some convenient reference frame.

(2) Rotate these co-ordinates through an angle (η) about a line through their origin and having direction numbers (λ, μ, ν).

(3) Transform the origin to an arbitrary point (x_0, y_0, z_0) in space.

(4) Minimize R or R_2 with respect to (x_0, y_0, z_0), (λ, μ, η). [ν is redundant and can be eliminated.]

It can be shown (Booth 1948) that if the frame of reference is triclinic such that

$$X\hat{O}Y = \gamma, \quad Y\hat{O}Z = \alpha, \quad Z\hat{O}X = \beta,$$

and if a (figure 6) is a line through O, having direction numbers (λ, μ, ν), then any point $P(x, y, z)$ becomes, after rotation through an angle η, in an anti-clockwise direction, the point $P'(\bar{x}, \bar{y}, \bar{z})$, where

$$\left.\begin{aligned}
\bar{x} &= x\cos\eta + \lambda(1-\cos\eta)\,[\lambda x + \mu y + \nu z + (\mu x + \lambda y)\cos\gamma + (\nu y + \mu z)\cos\alpha + (\lambda z + \nu x)\cos\beta] \\
&\qquad + \sin\eta[\{\gamma\alpha\beta\}(\mu x - \lambda y) + \{\alpha\}(\nu y - \mu z) + \{\alpha, \beta, \gamma\}(\lambda z - \nu x)], \\
\bar{y} &= y\cos\eta + \mu(1-\cos\eta)\,[\lambda x + \mu y + \nu z + (\mu x + \lambda y)\cos\gamma + (\nu y + \mu z)\cos\alpha \\
&\qquad + (\lambda z + \nu x)\cos\beta] + \sin\eta[\{\beta\gamma\alpha\}(\mu x - \lambda y) + \{\alpha\beta\gamma\}(\nu y - \mu z) + \{\beta\}(\lambda z - \nu x)], \\
\bar{z} &= z\cos\eta + \nu(1-\cos\eta)\,[\lambda x + \mu y + \nu z + (\mu x + \lambda y)\cos\gamma + (\nu y + \mu z)\cos\alpha \\
&\qquad + (\lambda z + \nu x)\cos\beta + \sin\eta[\{\nu\}(\mu x - \lambda y) + \{\gamma\alpha\beta\}(\nu y - \mu z) + \{\beta\gamma\alpha\}(\lambda z - \nu x)],
\end{aligned}\right\} \tag{10·1}$$

and

$$\{\alpha\} = \sin^2\alpha/D \quad \text{etc.},$$

$$\{\alpha\beta\gamma\} = (\cos\alpha\,\cos\beta - \cos\gamma)/D \quad \text{etc.},$$

$$D^2 = 1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\,\cos\beta\,\cos\gamma.$$

The direction numbers (λ, μ, ν) are related by

$$\lambda^2+\mu^2+\nu^2+2\lambda\mu\cos\gamma+2\mu\nu\cos\alpha+2\nu\lambda\cos\beta = 1. \tag{10·2}$$

Thus the whole transformation, including the translation of origin to (x_0, y_0, z_0), is given by

$$\bar{\bar{x}} = x_0+\bar{x}, \quad \bar{\bar{y}} = y_0+\bar{y}, \quad \bar{\bar{z}} = z_0+\bar{z}. \tag{10·3}$$

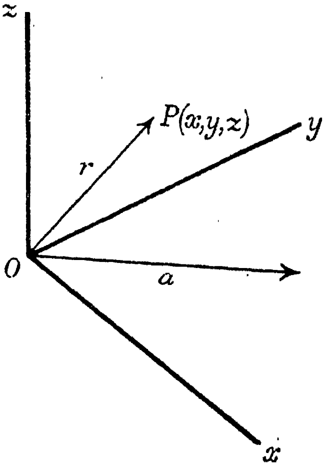


Figure 6

In the general, asymmetric case, typified by the definition of (9·2), a change of origin has no effect on the value of $F(h, k, l)$, so that (x_0, y_0, z_0) are redundant. When symmetry elements are present, however, this is no longer true. A little consideration of the operations of symmetry shows that, for the cases of reflexion and glide planes, and of centres and twofold axes, the effect on (9·2) is to make

$$\sum_{r=1}^{2N} f_r \sin\theta_r = A_1+A_2, \quad \sum_{r=1}^{2N} f_r \cos\theta_r = B_1+B_2, \tag{10·4}$$

where
$$A_m = \sum_{r=1}^{N} f_r \sin{}_m\theta_r, \quad B_m = \sum_{r=1}^{N} f_r \cos{}_m\theta_r \quad (m = 1, 2), \tag{10·5}$$

and all ${}_1\theta_r$ terms contain x_0 with a *positive* sign, whilst all ${}_2\theta_r$ terms have it similarly *negative*. The same applies to y_0 and z_0, but with possible differences in the terms included in the summations for A_m and B_m.

Different considerations apply when three-, four- or sixfold symmetry elements are present, but the treatment now to be given for the more elementary case will make the method of attack in these cases clear.

It may be worth while, at some future date, tabulating the particular results for each of the 230 space groups.

It is required to obtain the values of

$$\frac{\partial|F_c|}{\partial x_0}, \quad \frac{\partial|F_c|}{\partial y_0}, \quad \frac{\partial|F_c|}{\partial z_0}; \qquad \frac{\partial|F_c|}{\partial \eta}, \quad \frac{\partial|F_c|}{\partial \lambda}, \quad \frac{\partial|F_c|}{\partial \mu};$$

and of all second derivatives of the type

$$\frac{\partial^2|F_c|}{\partial a\,\partial b} \quad [a, b = x_0, y_0, z_0;\ \eta, \lambda, \mu].$$

From (9·2), (9·3), (10·3), (10·4) and (10·5) it follows that

$$\frac{\partial\,|F_c|}{\partial x_0} = 4\pi\frac{h}{a}[B_1A_2 - B_2A_1]/|F_c|, \tag{10·6}$$

with similar equations for $\partial\,|F_c|/\partial y_0$ and $\partial\,|F_c|/\partial z_0$. In a like manner it can be shown that

$$\frac{\partial^2|F_c|}{\partial x_0\,\partial y_0} = -16\pi^2\frac{hk}{ab}[(A_1A_2 + B_1B_2)\,|F_c|^2 + (B_1A_2 - B_2A_1)^2]/|F_c|^3, \tag{10·7}$$

from which the other derivatives can be derived by symmetry.

The derivatives with respect to the angles (η, λ, μ) are formally more complicated, and only the triclinic asymmetric case will be treated; others can be derived by identical methods.

First, from (9·2)

$$\frac{\partial\,|F_c|}{\partial \xi} = -\sum_r f_r \sin(\bar\theta_r - \alpha)\frac{\partial\bar\theta_r}{\partial\xi}, \tag{10·8}$$

where ξ is any one of the angle variables (η, λ, μ).

Now from (9·3) and (10·1)

$$\bar\theta_r = \theta_0 + \theta_r\cos\eta + \pi(1-\cos\eta)\left(\lambda\frac{h}{a} + \mu\frac{k}{b} + \nu\frac{l}{c}\right)[\lambda Q_{rx} + \mu Q_{ry} + \nu Q_{rz}]$$
$$+ \sin\eta[\lambda L_r + \mu M_r + \nu N_r],$$

where

$$Q_{rx} = \partial Q_r/\partial x_r,$$
$$Q_r = x_r^2 + y_r^2 + z_r^2 + 2x_ry_r\cos\gamma + 2y_rz_r\cos\alpha + 2z_rx_r\cos\beta,$$

and

$$L_r = -2\pi\left[\frac{h}{a}(\{\gamma\alpha\beta\}y_r - \{\alpha\beta\gamma\}z_r) + \frac{k}{b}(\{\beta\gamma\alpha\}y_r - \{\beta\}z_r) + \frac{l}{c}(\{\gamma\}y_r - \{\beta\gamma\alpha\}z_r)\right],$$
$$M_r = \text{etc.}, \tag{10·9}$$

whence

$$\frac{\partial\bar\theta_r}{\partial\eta} = -\theta_r\sin\eta + \pi\sin\eta\left(\lambda\frac{h}{a} + \mu\frac{k}{b} + \nu\frac{l}{c}\right)[\lambda Q_{rx} + \mu Q_{ry} + \nu Q_{rz}] + \cos\eta\,[\lambda L_r + \mu M_r + \nu N_r], \tag{10·10}$$

$$\frac{\partial\bar\theta_r}{\partial\lambda} = \pi(1-\cos\eta)\left\{\left(\frac{h}{a} + \frac{l}{c}\frac{\partial\nu}{\partial\lambda}\right)[\lambda Q_{rx} + \mu Q_{ry} + \nu Q_{rz}]\right.$$
$$\left. + \left(\lambda\frac{h}{a} + \mu\frac{y}{b} + \nu\frac{z}{c}\right)\left[Q_{rx} + Q_{rz}\frac{\partial\nu}{\partial\lambda}\right]\right\} + \sin\eta\left[L_r + N_r\frac{\partial\nu}{\partial\lambda}\right], \tag{10·11}$$

$$\frac{\partial\bar\theta_r}{\partial\mu} = \pi(1-\cos\eta)\left\{\left(\frac{k}{b} + \frac{l}{c}\frac{\partial\nu}{\partial\lambda}\right)[\lambda Q_{rx} + \mu Q_{ry} + \nu Q_{rz}]\right.$$
$$\left. + \left(\lambda\frac{h}{a} + \mu\frac{y}{b} + \nu\frac{z}{c}\right)\left[Q_{ry} + Q_{rz}\frac{\partial\nu}{\partial\mu}\right]\right\} + \sin\eta\left[M_r + N_r\frac{\partial\nu}{\partial\mu}\right], \tag{10·12}$$

and, from (10·2),

$$\frac{\partial\nu}{\partial\lambda} = -\frac{\lambda + \mu\cos\gamma + \nu\cos\beta}{\lambda\cos\beta + \mu\cos\alpha + \nu}, \tag{10·13}$$

$$\frac{\partial\nu}{\partial\mu} = -\frac{\lambda\cos\gamma + \mu + \nu\cos\alpha}{\lambda\cos\beta + \mu\cos\alpha + \nu}. \tag{10·14}$$

The second derivatives with respect to angles are obtained from (10·8)

$$\frac{\partial^2 |F_c|}{\partial\xi\,\partial\zeta} = \frac{1}{|F_c|}\left[\sum_r f_r \cos(\bar\theta_r - \alpha)\frac{\partial\bar\theta_r}{\partial\zeta}\right]\left[\sum_r f_r \cos(\bar\theta_r - \alpha)\frac{\partial\bar\theta_r}{\partial\xi}\right]$$
$$-\sum_r f_r \cos(\bar\theta_r - \alpha)\frac{\partial\bar\theta_r}{\partial\zeta}\frac{\partial\bar\theta_r}{\partial\xi} - \sum_r f_r \sin(\bar\theta_r - \alpha)\frac{\partial^2\bar\theta_r}{\partial\zeta\,\partial\xi}. \quad (10·15)$$

The only new $\bar\theta_r$ derivatives required to evaluate this are $\partial^2\bar\theta_r/\partial\zeta\,\partial\xi$, and these are readily obtained from (10·10) to (10·12):

$$\frac{\partial^2\bar\theta_r}{\partial\eta\,\partial\lambda} = \pi\sin\eta\left\{\left(\frac{h}{a}+\frac{l}{c}\frac{\partial\nu}{\partial\lambda}\right)[\lambda Q_{rx}+\mu Q_{ry}+\nu Q_{rz}]\right.$$
$$\left.+\left(\lambda\frac{h}{a}+\mu\frac{k}{b}+\nu\frac{l}{c}\right)\left[Q_{rx}+Q_{rz}\frac{\partial\nu}{\partial\lambda}\right]+\cos\eta\left(L_r+N_r\frac{\partial\nu}{\partial\lambda}\right)\right\}, \quad (10·16)$$

$$\frac{\partial^2\bar\theta_r}{\partial\eta\,\partial\mu} = \pi\sin\eta\left\{\left(\frac{k}{b}+\frac{l}{c}\frac{\partial\nu}{\partial\mu}\right)[\lambda Q_{rx}+\mu Q_{ry}+\nu Q_{rz}]\right.$$
$$\left.+\left(\lambda\frac{h}{a}+\mu\frac{k}{b}+\nu\frac{l}{c}\right)\left[Q_{ry}+Q_{rz}\frac{\partial\nu}{\partial\mu}\right]+\cos\eta\left[M_r+N_r\frac{\partial\nu}{\partial\mu}\right]\right\}, \quad (10·17)$$

$$\frac{\partial^2\bar\theta_r}{\partial\lambda\,\partial\mu} = \pi(1-\cos\eta)\left\{(\lambda Q_{rx}+\mu Q_{ry}+\nu Q_{rz})\frac{l}{c}\frac{\partial^2\nu}{\partial\lambda\,\partial\mu}\right.$$
$$+\left(\frac{h}{a}+\frac{l}{c}\frac{\partial\nu}{\partial\lambda}\right)\left[Q_{ry}+Q_{rz}\frac{\partial\nu}{\partial\mu}\right]+\left(\frac{k}{b}+\frac{l}{c}\frac{\partial\nu}{\partial\mu}\right)\left[Q_{rx}+Q_{rz}\frac{\partial\nu}{\partial\lambda}\right]$$
$$\left.+\left(\lambda\frac{h}{a}+\mu\frac{k}{b}+\nu\frac{l}{c}\right)Q_{rz}\frac{\partial^2\nu}{\partial\lambda\,\partial\mu}\right\}+N_r\frac{\partial^2\nu}{\partial\lambda\,\partial\mu}\sin\eta, \quad (10·18)$$

$$\frac{\partial^2\bar\theta_r}{\partial\eta^2} = -\theta_r\cos\eta+\pi\cos\eta\left(\lambda\frac{h}{a}+\mu\frac{k}{b}+\nu\frac{l}{c}\right)[\lambda Q_{rx}+\mu Q_{ry}+\nu Q_{rz}]$$
$$-\sin\eta[\lambda L_r+\mu M_r+\nu N_r], \quad (10·19)$$

$$\frac{\partial^2\bar\theta_r}{\partial\lambda^2} = \pi(1-\cos\eta)\left\{(\lambda Q_{rx}+\mu Q_{ry}+\nu Q_{rz})\frac{l}{c}\frac{\partial^2\nu}{\partial\lambda^2}\right.$$
$$\left.+2\left(\frac{h}{a}+\frac{l}{c}\frac{\partial\nu}{\partial\lambda}\right)\left[Q_{rx}+Q_{rz}\frac{\partial\nu}{\partial\lambda}\right]+\left(\lambda\frac{h}{a}+\mu\frac{k}{b}+\nu\frac{z}{c}\right)Q_{rz}\frac{\partial^2\nu}{\partial\lambda^2}\right\}+N_r\frac{\partial^2\nu}{\partial\lambda^2}\sin\eta, \quad (10·20)$$

$$\frac{\partial^2\bar\theta_r}{\partial\mu^2} = \pi(1-\cos\eta)\left\{(\lambda Q_{rx}+\mu Q_{ry}+\nu Q_{rz})\frac{l}{c}\frac{\partial^2\nu}{\partial\mu^2}\right.$$
$$\left.+2\left(\frac{k}{b}+\frac{l}{c}\frac{\partial\nu}{\partial\mu}\right)\left[Q_{ry}+Q_{rz}\frac{\partial\nu}{\partial\mu}\right]+\left(\lambda\frac{h}{a}+\mu\frac{k}{b}+\nu\frac{z}{c}\right)Q_{rz}\frac{\partial^2\nu}{\partial\mu^2}\right\}+N_r\frac{\partial^2\nu}{\partial\mu^2}\sin\eta, \quad (10·21)$$

where

$$\frac{\partial^2\nu}{\partial\lambda\,\partial\mu} = -\frac{(\lambda\cos\beta+\mu\cos\alpha+\nu)\left(\cos\gamma+\dfrac{\partial\nu}{\partial\mu}\cos\beta\right)-(\lambda+\mu\cos\gamma+\nu\cos\beta)\left(\cos\alpha+\dfrac{\partial\nu}{\partial\mu}\right)}{(\lambda\cos\beta+\mu\cos\alpha+\nu)^2}, \quad (10·22)$$

$$\frac{\partial^2\nu}{\partial\lambda^2} = -\frac{(\lambda\cos\beta+\mu\cos\alpha+\nu)\left(1+\dfrac{\partial\nu}{\partial\lambda}\cos\beta\right)-(\lambda+\mu\cos\gamma+\nu\cos\beta)\left(\cos\beta+\dfrac{\partial\nu}{\partial\lambda}\right)}{(\lambda\cos\beta+\mu\cos\alpha+\nu)^2}, \quad (10·23)$$

$$\frac{\partial^2\nu}{\partial\mu^2} = -\frac{(\lambda\cos\beta+\mu\cos\alpha+\nu)\left(1+\dfrac{\partial\nu}{\partial\mu}\cos\alpha\right)-(\lambda\cos\gamma+\mu+\nu\cos\alpha)\left(\cos\alpha+\dfrac{\partial\nu}{\partial\mu}\right)}{(\lambda\cos\beta+\mu\cos\alpha+\nu)^2}. \quad (10·24)$$

The direction and magnitude of the steepest descent are now obtained by the straightforward application of either (7·3) or (7·5).

It may be argued that the complexity of the expressions, given above, is such that computation is not practicable. This is not true; an examination shows that many of the component parts are identical, and this, taken in conjunction with the fact that the solution converges to its final value in 3-4 iterations, makes the total time required for a determination less than that taken for a single complete set of three-dimensional Fourier sections.

In the case of solution by an electronic computer, the whole process is very simple, and the machine will continue refinement until a solution of preassigned accuracy is reached.

A particular advantage of the method lies in the fact that the $3N$ atomic co-ordinates (x_r, y_r, z_r) do not have to be adjusted during the refinement, the only corrections being to the six quantities (x_0, y_0, z_0), (η, λ, μ).

11. THE METHOD OF CONFIGURATIONAL EQUATIONS

Should it be desired to make direct use of the configurational equations

$$\begin{aligned} c_{st} \equiv (x_s-x_t)^2+(y_s-y_t)^2+(z_s-z_t)^2+2(x_s-x_t)(y_s-y_t)\cos\gamma \\ +2(y_s-y_t)(z_s-z_t)\cos\alpha+2(z_s-z_t)(x_s-x_t)\cos\beta-l_{st}^2 = 0, \quad (11{\cdot}1)\end{aligned}$$

in order to insure minimization to a possible configuration, the formal technique is particularly simple. Instead of minimizing R or R_2, as defined previously, new functions

$$c \equiv R+\sum_{st}\lambda_{st}\,|\,c_{st}\,| \qquad (11{\cdot}2)$$

or

$$c_2 \equiv R_2+\sum_{st}\lambda_{st}c_{st}^2 \qquad (11{\cdot}3)$$

are used, the λ_{st} being positive *weighting* factors.

It is not necessary to include all c_{st} in the summation over (s, t), but only those sufficient to define uniquely the molecular configuration. Thus, for a rigid molecule, any three non-collinear atoms might be chosen as 'standards', and the position of any other atom defined by its distances from these three. The summation would then be written

$$\sum_{st}\lambda_{st}e_{st} = \lambda_{12}e_{12}+\lambda_{13}e_{13}+\lambda_{23}e_{23}+\sum_{t=4}^{N}(\lambda_{1,t}e_{1,t}+\lambda_{2,t}e_{2,t}+\lambda_{3,t}e_{3,t}), \qquad (11{\cdot}4)$$

where e_{st} represents either $|\,c_{st}\,|$ or c_{st}^2 according as C or C_2 is to be minimized.

The partial derivatives required for the steepest descents process can now be written down:

$$\begin{aligned} \frac{\partial c}{\partial x_{rj}} &= \frac{\partial R}{\partial x_{rj}}+\sum_{p=1}^{3}\lambda_{pr}\frac{c_{pr}}{|\,c_{pr}\,|}\frac{\partial c_{pr}}{\partial x_{rj}} \quad (r \neq 1,2,3), \\ &= \frac{\partial R}{\partial x_{rj}}+\sum_{t=1}^{N}{}'\lambda_{rt}\frac{c_{rt}}{|\,c_{rt}\,|}\frac{\partial c_{rt}}{\partial x_{rj}} \quad (r = 1,2,3), \qquad (11{\cdot}5)\end{aligned}$$

$$\begin{aligned} \frac{\partial c_2}{\partial x_{rj}} &= \frac{\partial R_2}{\partial x_{rj}}+2\sum_{p=1}^{3}\lambda_{pr}c_{pr}\frac{\partial c_{pr}}{\partial x_{rj}} \quad (r \neq 1,2,3), \\ &= \frac{\partial R_2}{\partial x_{rj}}+2\sum_{t=1}^{N}{}'\lambda_{rt}c_{rt}\frac{\partial c_{rt}}{\partial x_{rj}} \quad (r = 1,2,3), \qquad (11{\cdot}6)\end{aligned}$$

where the prime (′) indicates that the term for $t = r$ is to be omitted, and, for example,

$$\frac{\partial c_{rt}}{\partial x_{r1}} = 2\{(x_r - x_t) + (y_r - y_t)\cos\gamma + (z_r - z_t)\cos\beta\}, \tag{11·7}$$

which follows from (11·1) and the definition of x_{rj} given in (9·6). The derivatives of R and R_2 are obtained from (9·8) to (9·11).

Second derivatives are obtained in a similar manner and are

$$\begin{aligned}\frac{\partial^2 c}{\partial x_{rj}\partial x_{r'j'}} &= \frac{\partial^2 R}{\partial x_{rj}\partial x_{r'j'}} + \sum_{p=1}^{3} \lambda_{pr} \frac{c_{pr}}{|c_{pr}|} \frac{\partial^2 c_{pr}}{\partial x_{rj}\partial x_{r'j'}} \quad (r \neq 1, 2, 3)\\ &= \frac{\partial^2 R}{\partial x_{rj}\partial x_{r'j'}} + \sum_{t=1}^{N}{}' \lambda_{rt} \frac{c_{rt}}{|c_{rt}|} \frac{\partial^2 c_{rt}}{\partial x_{rj}\partial x_{r'j'}} \quad (r = 1, 2, 3),\end{aligned} \tag{11·8}$$

$$\begin{aligned}\frac{\partial^2 c_2}{\partial x_{rj}\partial x_{r'j'}} &= \frac{\partial^2 R_2}{\partial x_{rj}\partial x_{r'j'}} + 2\sum_{p=1}^{3} \lambda_{pr} \left[\frac{\partial c_{pr}}{\partial x_{rj}} \frac{\partial c_{pr}}{\partial x_{r'j'}} + c_{pr} \frac{\partial^2 c_{pr}}{\partial x_{rj}\partial x_{r'j'}}\right] \quad (r \neq 1, 2, 3)\\ &= \frac{\partial^2 R_2}{\partial x_{rj}\partial x_{r'j'}} + 2\sum_{t=1}^{N}{}' \lambda_{rt} \left[\frac{\partial c_{rt}}{\partial x_{rj}} \frac{\partial c_{rt}}{\partial x_{r'j'}} + c_{rt} \frac{\partial^2 c_{rt}}{\partial x_{rj}\partial x_{r'j'}}\right] \quad (r = 1, 2, 3).\end{aligned} \tag{11·9}$$

It will be seen that

$$\frac{\partial c_{rt}}{\partial x_{r'j'}} = 0 \quad (r' \neq r \text{ or } t), \tag{11·10}$$

and when $r' = r$ or t, $\partial c_{rt}/\partial x_{r'j'}$ is given by expressions of the type (11·7). The other derivatives are

$$\frac{\partial^2 c_{rt}}{\partial x_{rj}\partial x_{r'j'}} = 0 \quad (r' \neq r \text{ or } t), \tag{11·11}$$

or if $r' = r$ or t, they are obtained by differentiating expressions of type (11·7), so that, for example

$$\frac{\partial^2 c_{rt}}{\partial x_{r1}\partial x_{t2}} = -2\cos\gamma. \tag{11·12}$$

The choice of the weighting factors λ_{st} is somewhat complicated. If the configuration of the structure is fairly certain it will be best to make all the λ_{st} equal to some fixed positive number λ whose magnitude is varied as the determination proceeds. Thus at the start, λ will be made sufficiently large to ensure that the configurational equations exert a predominating effect on the minimization, whereas towards the end of the refinement, when the unique solution is well established, it will be reduced to zero so that the individual details of bonding will be obtained.

Although the treatment of this section appears to be formally simpler than that of § 10, the application is more laborious. This is because *all* the corrections for *all* the atoms have to be calculated at each stage, and the summations in (7·3) and (7·5) become very long. In addition, the choice of an initial value for λ is critical, since too small a value may lead to a configuration radically at variance with that postulated, whereas too large a value may cause the refinement to be protracted.

In the case of a molecule consisting of several rigid parts connected together in a non-rigid manner (e.g. with free rotation about some bond), an exactly similar treatment can be given. All that is required is the breaking up of the left-hand side of (11·4) into parts relating to each molecular segment, together with suitable connexions expressing the relation of one part to the next.

12. LARGE MOLECULES

In the case of fairly large molecules (*c.* 100 atoms other than hydrogen) few cases will arise when the structure can be considered rigid, so that the method of § 10 cannot be used directly. Almost all molecules will, however, contain groupings which are rigid, in the mechanical sense, and in these cases a slight modification of § 10 enables an analysis to be made.

All that has to be done is to define the position of each group of atoms by co-ordinates and angles, and then to examine the whole system by the methods of §§ 10 and 11.

The connexity of the various groups is introduced by the following procedure. Suppose atom 1 of group g is connected to atom r of group G, then the co-ordinates (x_0, y_0, z_0) for the group g are simply taken as (x_r, y_r, z_r)—i.e. as the co-ordinates of the rth atom of group G. The analysis follows closely that of §§ 10 and 11 and will not be reproduced here.

13. REFINEMENT OF DENSITY DISTRIBUTIONS

In the case of mega-molecular substances none of the above special methods will be directly applicable owing to the lack of knowledge of either atomic positions or configurations. The general method of steepest descents is, however, still applicable when any details whatsoever regarding the general shape of the molecule are available.

Suppose that considerations of molecular weight, space group, and packing lead to the supposition that a particular mega-molecule has a certain density distribution in space. It will now be shown how this distribution can be refined so as to minimize R_2 even although no detailed atomic positions are known.

Let $\rho(x, y, z)$ be the density distribution assumed normalized, so that

$$\int_V \rho(x, y, z)\, dV = \sum_r N_r, \tag{13·1}$$

where V is the volume of the whole unit cell and $\sum_r N_r$ represents the total number of electrons in this unit. The structure factor, $F_c(h, k, l)$, is defined by

$$F_c(h, k, l) = \int_V \rho(x, y, z)\, e^{i\theta}\, dV. \tag{13·2}$$

Now if a certain density distribution is postulated, the only method of defining it when at all complex is to give its values at points on a three-dimensional lattice. In this case (13·1) and (13·2) can be replaced by

$$\sum_p \rho(p) = \sum_r N_r, \tag{13·3}$$

$$F_c(h, k, l) = \sum_p \rho(p)\, e^{i\theta_p}, \tag{13·4}$$

θ_p being defined by $2\pi(hx_p/a + ky_p/b + lz_p/c)$, where p is the particular lattice point.

It follows from (13·4) that

$$|F_c(h, k, l)|^2 = [\sum_p \rho(p) \cos\theta_p]^2 + [\sum_p \rho(p) \sin\theta_p]^2. \tag{13·5}$$

The minimization problem is exactly the same as that of (7) with the exception that the variables are now the $\rho(p)$. Thus (7·5) becomes

$$\epsilon_p = -\left[\sum_p \left(\frac{\partial\phi}{\partial\rho(p)}\right)^2\right]\frac{\partial\phi}{\partial\rho(p)} \Big/ \sum_{p,t} \frac{\partial^2\phi}{\partial\rho(p)\,\partial\rho(t)}\,\frac{\partial\phi}{\partial\rho(p)}\,\frac{\partial\phi}{\partial\rho(t)}, \tag{13·6}$$

where ϵ_p is now the correction to $\rho(p)$.

Now any variation in one of the $\rho(p)$ must be such that (13·3) is satisfied; that is one of the $\rho(p)$, say $\rho(p_0)$, is a dependent variable and consequently

$$\frac{\partial\rho(p_0)}{\partial\rho(p)} = -1. \tag{13·7}$$

The investigation of the derivatives of ϕ in the cases of R and R_2 follows closely that of § 9, and it is seen that

$$\frac{\partial R}{\partial\rho(p)} = \sum_{h,k,l} \frac{|F_o| - |F_c|}{||F_o| - |F_c||}\,\frac{\partial|F_c|}{\partial\rho(p)}, \tag{13·8}$$

$$\frac{\partial^2 R}{\partial\rho(p)\,\partial\rho(t)} = \sum_{h,k,l} \frac{|F_o| - |F_c|}{||F_o| - |F_c||}\,\frac{\partial^2|F_c|}{\partial\rho(p)\,\partial\rho(t)}. \tag{13·9}$$

The preliminary forms of (9·10) and (9·11) are still the same and it is seen that for both R and R_2 the derivatives

$$\frac{\partial|F_c|}{\partial\rho(p)} \quad\text{and}\quad \frac{\partial^2|F_c|}{\partial\rho(p)\,\partial\rho(t)}$$

are required; they are at once obtained from (13·5) and (13·7)

$$\frac{\partial|F_c|}{\partial\rho(p)} = \cos(\theta_p - \alpha) - \cos(\theta_{p_0} - \alpha), \tag{13·10}$$

$$\begin{aligned}\frac{\partial^2|F_c|}{\partial\rho(p)\,\partial\rho(t)} = \frac{1}{|F_c|}\{&1 + \cos(\theta_t - \theta_p) - \cos(\theta_t - \theta_{p_0}) - \cos(\theta_p - \theta_{p_0})\\ &- [\cos(\theta_p - \alpha) - \cos(\theta_{p_0} - \alpha)]\,[\cos(\theta_t - \alpha) - \cos(\theta_{p_0} - \alpha)]\}.\end{aligned} \tag{13·11}$$

Thus (13·6), (13·8), (13·9), (9·10), (9·11), (13·10) and (13·11) contain the solution of the minimization problem for the case in which the molecular shape is given as a density distribution at lattice points.

This variant of the method of 'steepest descents' is particularly easy to apply, since all the trigonometric functions, being defined at points on a predetermined lattice, are invariant throughout a set of refinements.

There is one disadvantage of the lattice-point procedure. To define adequately a large molecule a reasonable fineness of mesh is required; now even a coarse net of 10 points per cell edge involves 10^3 lattice points, and an increase to 100 divisions increases this number to 10^6. Even the latter mesh is rather coarse for the examination of biomolecular substances, and the work of computation is of an order which would be straining the resources even of an electronic computer. The situation is not, however, quite as gloomy as might first appear; the treatment given applies to a tri-clinic asymmetric space group and, as symmetry elements are introduced, the proportion of the unit cell which must be examined decreases, consequently in any real example not more than $2\cdot5 \times 10^5$ points are likely to be required, and this number is more reasonable for the types of computer at present under construction.

14. PHASE INDEPENDENT VARIANT

An interesting variant of all the above methods becomes available when there is a sufficient number of planes of observed intensity zero. (This is excluding planes zero because of symmetry elements.) In this case the simple expressions

$$O = \sum_{F_o=0} |F_c| \tag{14·1}$$

and

$$O_2 = \sum_{F_o=0} |F_c|^2 \tag{14·2}$$

can be minimized, and all the differential coefficients are much simplified.

15. WEIGHTING FACTORS

A considerable advantage of the steepest descent technique over the standard Fourier methods lies in the fact that experimental data can be used in weighted form. Thus, if certain planes are suspected to suffer appreciable extinction they can be assigned low weights in the calculation. The quantity to be minimized, when weights are included, is simply

$$W = \sum_{h,k,l} w(h,k,l)\, ||F_o| - |F_c||, \tag{15·1}$$

or

$$W_2 = \sum_{h,k,l} w(h,k,l)\, (|F_o|^2 - |F_c|^2)^2. \tag{15·2}$$

The question of a systematic choice of the weights, $w(h,k,l)$, is more speculative. Work on simple structures (Booth 1946*c*) appears to show that the probable experimental error in any $F_0(h,k,l)$ is independent of the magnitude of that $F_0(h,k,l)$. This would suggest that the assumption

$$w(h,k,l) \propto F_0(h,k,l) \tag{15·3}$$

is appropriate. Much more extensive investigation is needed, however, before a definite assertion can be made, especially in the case of macro-molecular crystals.

16. THE SCATTERING FACTORS

It has been suggested by Hughes (1941), in connexion with the least squares refinement, that a lack of knowledge of the *absolute* values of $|F_{\text{obs.}}|$ and the uncertainty of the values of f—the atomic scattering factors—might make a minimization based on R or R_2 unsatisfactory. The steepest descents technique does **not** suffer this disadvantage.

In the case of non-absolute $|F_{\text{obs.}}|$ values it suffices to assume $|F_{\text{abs.}}| = K\,|F_{\text{obs.}}|$ in (7·3), (7·5) and (9·7). K is then varied in exactly the same manner as the x_{rj} and corrections, ϵ_k, to it obtained as for the former. The required derivatives are very simple:

$$\frac{\partial R}{\partial K} = \sum_{h,k,l} \frac{K|F_o|-|F_c|}{|K|F_o|-|F_c||}|F_o|, \tag{16·1}$$

$$\frac{\partial R_2}{\partial K} = \sum_{h,k,l} 4|F_o|^2 K(K^2|F_o|^2-|F_c|^2). \tag{16·2}$$

For atomic scattering factors the same sort of process applies thus, by assuming that the true factor for the rth atom is given by

$$f_{tr} = f_r \, e^{-B_r\rho^2}, \tag{16·3}$$

where ρ is the reciprocal spacing of the plane (h,k,l), B_r is a constant for atom r, and f_r is the tabular scattering factor. It is quite simple to minimize R or R_2, not only with respect to the atomic parameters, but also with respect to the B_r.

In most cases it will be sufficient to take a common value B for B_r to cover all the atoms, so that the only change in the previous treatment is the inclusion of a correction ϵ_B and of terms $(\partial\phi/\partial B)$, $(\partial^2\phi/\partial B^2)$ and $(\partial^2\phi/\partial B\,\partial x_{rj})$ in (7·3) and (7·5). F_c now becomes $F_c e^{-B\rho^2}$, and the new derivatives required for (9·8), (9·9), (9·10) and (9·11) are simply

$$\frac{\partial|F_c|}{\partial B} = -\rho^2 e^{-B\rho^2}|F_c|, \tag{16·4}$$

$$\frac{\partial^2|F_c|}{\partial B^2} = \rho^4 e^{-B\rho^2}|F_c|, \tag{16·5}$$

$$\frac{\partial^2|F_c|}{\partial B\,\partial x_{rj}} = -\rho^2 e^{-B\rho^2}\frac{\partial|F_c|}{\partial x_{rj}}. \tag{16·6}$$

The treatment for more general constants B_r is similar and will not be given here.

17. STERIC FACTORS

A question which has often been posed is that of determining a given structure using only the X-ray intensities and the known fact that there is a certain minimum distance of approach of atoms. Although this type of data can be twisted into a form to which the method of configurational equations of § 11 may be applied, a somewhat simpler procedure is available which possesses the added advantage enabling the idea of one atom being *between* a_1 and a_2 Å $(a_2 > a_1)$ from another to be used.

The general scheme is to add to the function R_2 (or R or O and O_2) another function which becomes large (or even infinite) at the two limits and has a minimum between.

A convenient function for this purpose is

$$s = 1/(d_u^2-d^2)(d^2-d_l^2), \tag{17·1}$$

where d_u is the upper limit of separation, d is the actual separation, and d_l is the least possible separation. Figure 7 gives a rough illustration of the form of s.

To make use of the technique, instead of minimizing R or R_2, the functions

$$S = R + \sum_{p \neq q} s_{pq}, \tag{17·2}$$

or

$$S_2 = R_2 + \sum_{p \neq q} s_{pq}, \tag{17·3}$$

are used and the derivatives required become

$$\frac{\partial S}{\partial x_{sj}} = \frac{\partial R}{\partial x_{sj}} + \sum_{p \neq q} \frac{(2d_{pq}^2 - d_l^2 - d_u^2)}{(d_u^2 - d_{pq}^2)(d_{pq}^2 - d_l^2)} \frac{\partial d_{pq}^2}{\partial x_{sj}}. \tag{17·4}$$

It should be noted that $\partial d_{pq}^2/\partial x_{sj} = 0$ unless $s = p$ or q. Its value when not zero is found from (11·1) to be

$$\frac{\partial d_{pq}^2}{\partial x_{p1}} = \frac{\partial d_{pq}^2}{\partial x_{q1}} = 2(x_p - x_q) + 2(y_p - y_q)\cos\gamma + 2(z_p - z_q)\cos\beta, \tag{17·5}$$

$$\frac{\partial d_{pq}^2}{\partial x_{p2}} = \frac{\partial d_{pq}^2}{\partial x_{q2}} = 2(y_p - y_q) + 2(x_p - x_q)\cos\gamma + 2(z_p - z_q)\cos\alpha, \tag{17·6}$$

$$\frac{\partial d_{pq}^2}{\partial x_{p3}} = \frac{\partial d_{pq}^2}{\partial x_{q3}} = 2(z_p - z_q) + 2(y_p - y_q)\cos\alpha + 2(x_p - x_q)\cos\beta. \tag{17·7}$$

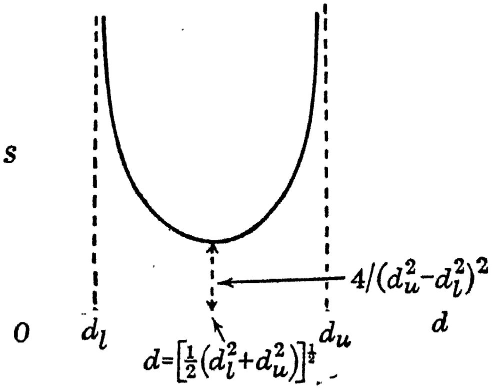


Figure 7

The corrections ϵ_{rj} are still given by (7·3) or (7·5), where the substitutions

$$_0R_{(2)} = {_0S_{(2)}}, \quad \frac{\partial \phi}{\partial x_{rj}} = \frac{\partial S_{(2)}}{\partial x_{rj}}, \quad \frac{\partial^2 \phi}{\partial x_{rj}\partial x_{r'j'}} = \frac{\partial^2 S_{(2)}}{\partial x_{rj}\partial x_{r'j'}}$$

are made.

References

Booth, A. D. 1945 *Trans. Faraday Soc.* **41**, 434.
Booth, A. D. 1946*a* *Trans. Faraday Soc.* **42**, 444.
Booth, A. D. 1946*b* *Trans. Faraday Soc.* **42**, 617.
Booth, A. D. 1946*c* *Proc. Roy. Soc.* A, **188**, 77.
Booth, A. D. 1948 *Math. Gaz.* **32**, 70.
Bragg, W. L. 1929 *Proc. Roy. Soc.* A, **123**, 537.
Hughes, E. W. 1941 *J. Amer. Chem. Soc.* **63**, 1737.

Studies in polymerization

III. The polymerization of methyl methacrylate

By C. H. Bamford and M. J. S. Dewar

Courtaulds Ltd. Research Laboratory, Maidenhead, Berks.

(*Communicated by A. H. Wilson, F.R.S.—Received* 29 *June* 1948—
Revised 15 *November* 1948—*Read* 10 *February* 1949)

The polymerization of methyl methacrylate has been studied by the viscosity method (part I, 1948). The reaction is shown to be a normal radical chain process with velocity constants at 0° C given in table 3.

It has been found that irradiation of methyl methacrylate by ultra-violet light produces a catalyst, while an inhibitor is formed by a thermal reaction. These facts account for the difficulties experienced by other workers. A reinvestigation of the vapour-phase polymerization shows that they also account for the anomalous polymerization observed by Melville (1937), and attributed by him to a 'bound-bond' mechanism.

Chain termination in the polymerization of styrene and methyl methacrylate is shown to occur by disproportionation.

In part I of this series (Bamford & Dewar 1948*a*) we described a method for determining the absolute velocity constants in vinyl polymerizations, and showed that it could be applied successfully to styrene. Here we shall describe its application to methyl methacrylate. This monomer was chosen partly because of the technical importance of the polymer, and partly because the work of a large number of investigators has shown that its general behaviour is rather similar to that of styrene.* Further, Melville (1937) has claimed that the photopolymerization of the vapour follows a special course, involving unusual intermediates. It was of interest to see whether the liquid-phase polymerization showed any such peculiar features.

We soon found that the reaction presents a number of anomalies which were not encountered with styrene, and which had to be resolved before any quantitative measurements could be made. The first part of this paper describes a solution of this problem, and the second part the evaluation of the constants.

EXPERIMENTAL METHODS

(*a*) *Purification of methyl methacrylate*

The monomer, purchased from I.C.I., was dried over calcium chloride, and submitted to a preliminary distillation in an inert atmosphere through a 50-plate column. The subsequent operations were carried out entirely *in vacuo* ($p < 10^{-4}$ mm.). After the monomer had been thoroughly degassed, the purification involved treatment with lithium, heat and ultra-violet light. In preliminary work the order of the stages was varied; in the final purification for the determination of velocity constants the following procedure was found most satisfactory. The ester was first heated for some months at 100° C until it had polymerized to a syrup. This treatment removes any traces of normal catalysts and inhibitors—peroxides, hydroquinone, etc. After several fractional distillations, the methacrylate was treated overnight with clean lithium. Two further distillations followed, and the ester was then irradiated at 0° C in a quartz reaction vessel with the light of a hot mercury arc filtered through Pyrex, until considerable polymerization had occurred. The final distillation into viscometers was carried out in a system built entirely of quartz, without taps or greased joints, and previously degassed by heating to bright redness *in vacuo*.

---

* The most detailed investigation is that of Schulz & Blaschke (1942) and Schulz & Harborth (1947). See also Norrish & Brookman (1939), Norrish & Smith (1942), Trommsdorff (1944), Baxendale, Evans & Kilham (1946), Walling & Briggs (1946) and Melville (1937).

Some experiments were carried out with a specimen of monomer obtained from I.G. Farbenindustrie, with similar results.

(*b*) *Measurement of rates of polymerization*

The polymerization was followed as in part I by the rate of change of viscosity. The viscometers, of fused silica, differed somewhat from those described in part I. The appendage (*A*, figure 1) was attached by a long tube without any constriction. When polymerization had proceeded too far for further measurements to be practicable, the liquid was poured into the appendage and distilled back into the visco-

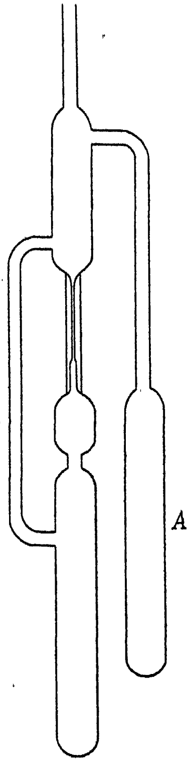


FIGURE 1. Quartz vacuum viscometer.

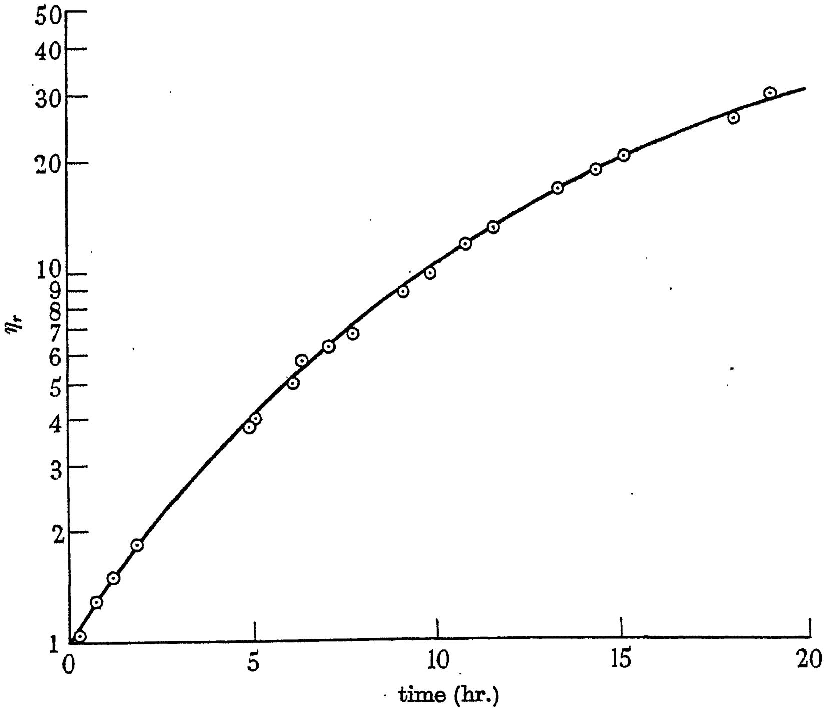


FIGURE 2. Spontaneous retardation of polymerization.

meter; the residual polymer was then so situated that it was unnecessary to seal off the appendage, since the polymer never came into contact with the monomer. The recovery of monomer could thus be repeated many times, and a far greater number of measurements could be made with one specimen. With viscometers of this type the times of flow (approximately 80 sec. initially) were reproducible to ± 0·05 sec. and were therefore measured with $\frac{1}{100}$ sec. stop-watches.

The observed specific viscosities were corrected to ideal specific viscosities,* as in part I, by using the relation of Schulz & Sing (1943), which we found to hold accurately for our polymers up to $\eta_{\text{sp.}} = 2{\cdot}5$.

In other respects the experimental method followed closely that described in part I.

MECHANISM OF POLYMERIZATION

We could not obtain reproducible thermal viscosity rates with the technique used previously for styrene. The variations were very large, rates at 25° C varying from 10^{-4} to less than 10^{-8} sec.$^{-1}$. If the methacrylate had been irradiated immediately before distillation into the viscometers, high rates were usually obtained, but if heating preceded distillation, little or no reaction could be observed. This result was surprising because the previous treatment should have removed all normal catalysts and inhibitors. Moreover, if a specimen of methacrylate which was polymerizing rapidly (in the dark) at 25° C was heated to 100° C for a few minutes, and then cooled to 25° C, the polymerization virtually ceased; a short irradiation was then sufficient to start the (dark) reaction once more. Apparently heating must either destroy a catalyst formed by irradiation, or produce an inhibitor which can be removed by irradiation.

Furthermore, after the light is cut off, the rate of polymerization steadily diminishes with time. Figure 2 shows this effect for a sample of monomer which had been strongly irradiated, and was polymerizing rapidly initially. Since it was necessary to follow the reaction to very high specific viscosities where the Schulz-Sing relation does not hold, we have followed Foord (1940) in plotting $\log \eta_r$ against time. If $d\eta/dt$ were constant, the $\log \eta_r$ against t plot would be approximately linear. Figure 3 shows the effect for a sample which contained initially less catalyst or more inhibitors; in this case the final rate was extremely small.

The gradual decay of a catalyst, or accumulation of an inhibitor, would account for these results. Experiments were therefore carried out to distinguish between these possibilities.

(i) Polymerization could be arrested not only by heating the liquid, but also by heating the vapour above the monomer in the viscometer to 180° C for a few minutes, and then cooling and shaking. This result can only be explained if heating the vapour produces an inhibitor.

(ii) Similarly, the rate of polymerization could be greatly increased by irradiating the vapour only with 2537 Å. This must imply that irradiation produces a catalyst of some kind.

---

* The ideal specific viscosity η is defined by $\eta = c[\eta]$, where c is the base molar concentration, and $[\eta]$ the intrinsic viscosity of the polymer. Throughout this paper η will signify the ideal specific viscosity.

(iii) After irradiation of the liquid at 0° C with 2537 Å, the rate, initially high, fell after some time to a constant and reproducible value. This is shown in figure 4. Over very long periods the rate showed a further gradual falling off, presumably due to accumulation of inhibitor (see above). The initial much more rapid fall must be due to a different cause, most probably the decay of a catalyst produced photochemically. The reproducible intermediate rate then corresponds to the true thermal rate.

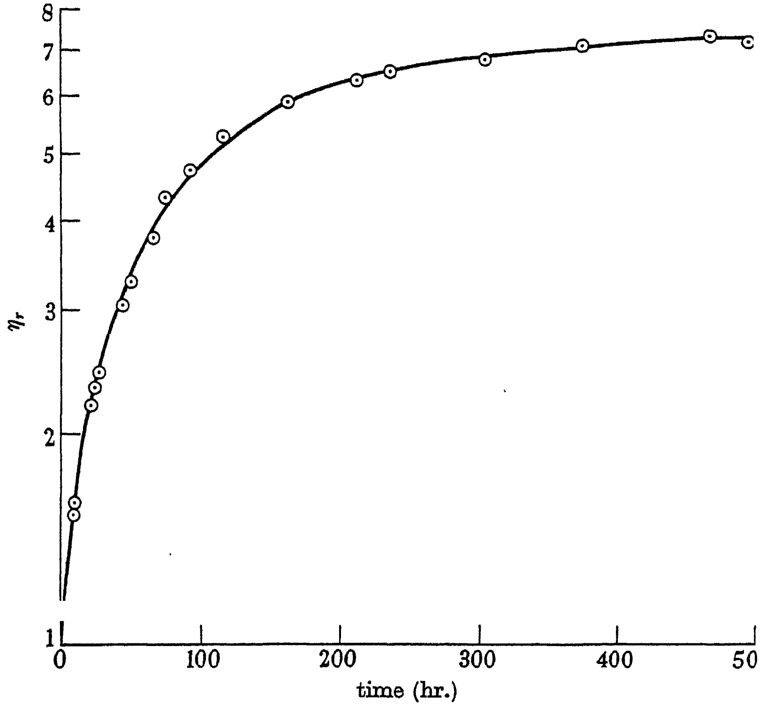


FIGURE 3. Spontaneous retardation of polymerization.

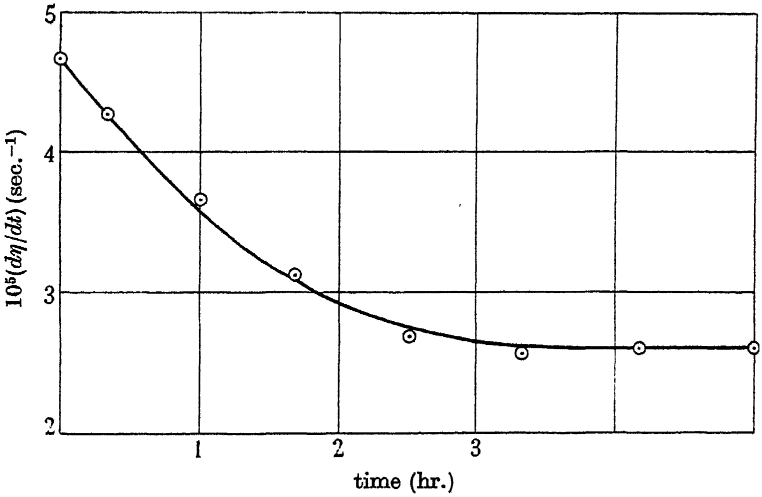


FIGURE 4. Decay of photocatalyst at 0° C.

These experiments show that both phenomena envisaged above do in fact take place.

Walling & Briggs (1946) studied the polymerization of methyl methacrylate at comparatively high temperatures (100 to 150° C) using monomer which had been heated for long periods at 100° C, and then carefully protected from light. Their material must have contained much inhibitor, and their rates could only have been reproducible if the concentration of inhibitor reached a stationary value during the heating. Walling & Briggs deduced from their experiments the rate of initiation and concluded that it had a low activation energy, and a frequency factor of approximately 1. It is, however, evident that the quantity measured by Walling & Briggs was not the true rate of formation of active centres, and the very low frequency factor is not significant. The thermal initiation reaction with styrene has a normal frequency factor (see part I).

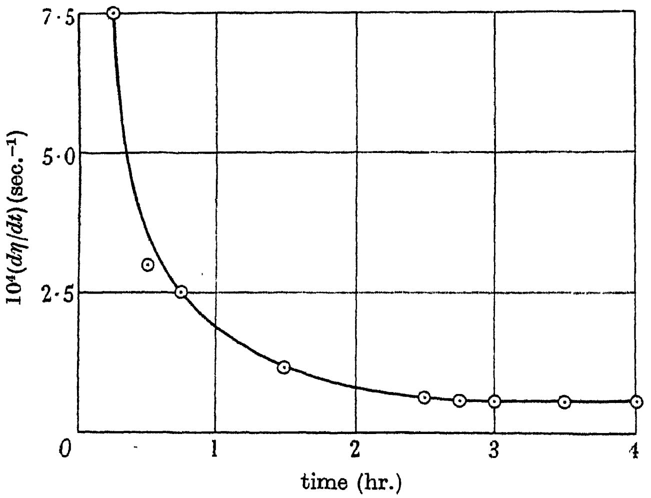


FIGURE 5. Formation of inhibitor at 100° C.

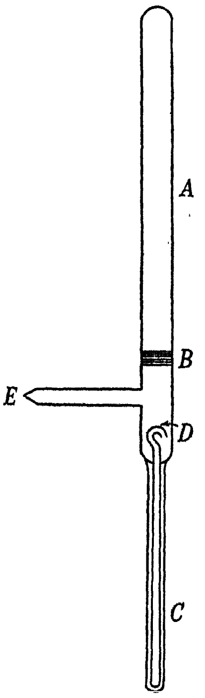


FIGURE 6. Apparatus for vapour-phase experiments.

In figure 5 the viscosity rate at 100° C is plotted against time of heating at 100° C; the rate was determined by periodically cooling the viscometer to 25° C and measuring the viscosity at that temperature. The rate falls very rapidly at first but attains a constant value $5{\cdot}5 \times 10^{-5}$ sec.$^{-1}$ after about $2\frac{1}{2}$ hr. This constant rate was quite reproducible and is less than one-third of the true thermal rate at 25° C ($2{\cdot}0 \times 10^{-4}$ sec.$^{-1}$). The true thermal rate at 100° C must be very much greater than this.

THE 'DOUBLE-BOND MECHANISM'

Melville (1937) observed that in the photopolymerization of methyl methacrylate vapour, reaction continued in the dark after the light had been cut off. He ascribed this 'photochemical after-effect' to a special kind of polymerization in which the active centres are not radicals but contain 'activated double bonds', and in which there is no termination process, so that centres, once formed, continue to grow indefinitely. Indeed, Melville (1939, 1940, 1941) has claimed that molecules 1 cm. long can be obtained by continued growth under suitable conditions.

The phenomenon reported by Melville is evidently similar to the one observed by us in the liquid-phase polymerization; but in our case at least his mechanism cannot apply. Thus:

(*a*) Melville's mechanism requires the rate of polymerization to remain constant while the molecular weight of the polymer steadily increases. The viscosity rate would then increase with time, in direct contradiction to our results.

(*b*) If the liquid monomer is irradiated at 0° C in the appendage of a viscometer, and then distilled into the viscometer while the appendage is maintained at 0° C, the initial rate of polymerization at 0° C is very high (up to five times the thermal rate). Evidently the active substance produced by irradiation is volatile at 0° C; it cannot therefore consist of growing polymer chains, but must be some compound of low molecular weight.

(*c*) The reaction can be completely suppressed by adding traces of standard radical chain inhibitors, e.g. quinone. This suggests very strongly that the intermediates are free radicals.

The mechanism suggested by Melville is in any case open to theoretical objections, and could only be accepted if it were based on irrefutable experimental evidence. It is difficult to see how the lifetime of an excited state of a double bond could extend to several hours or even days, even if transition to the ground state were forbidden. Since his suggestion has been widely discussed, however, we thought it desirable that the work should be repeated.

VAPOUR-PHASE EXPERIMENTS

The first series of experiments was carried out with an arrangement similar to that used by Melville (1937). The apparatus is shown in figure 6. A quartz reaction tube *A*, with total wall area 25 cm.^{2}, was attached by the graded seal *B* to a soda-glass capillary *C*. A tube *D* prevented any particles from the vapour falling into the capillary. Pure monomer was distilled into *C in vacuo*, through the constriction at *E*, which was then sealed. *A* and *C* were separately jacketed, the former in silica, so that they could be maintained at different temperatures. Usually *A* was kept at 25° C, and *C* at a lower temperature. The disappearance of monomer from *C* was followed by reading the level of the liquid to 0·01 mm. with a cathetometer. Tube *A* could be irradiated with 2537 Å. The following points were soon established:

(i) After irradiation, the rate of dark polymerization was very variable, being higher the larger the amount of polymer deposited in *A*. Although Melville claimed that after the 'induction period' the rates were constant, it is clear from his paper that they did in fact vary considerably (see Melville 1937, for example, table IV and p. 527).*

(ii) Even after prolonged periods of irradiation the dark rate fell off steadily with time, although the variation was less the greater the quantity of polymer deposited in *A*. Melville (1937) claimed that when the amount of polymer in the reaction vessel exceeded about 10^{19} molecules/cm.^{2}, the rate of the dark reaction remained constant for long periods of time.

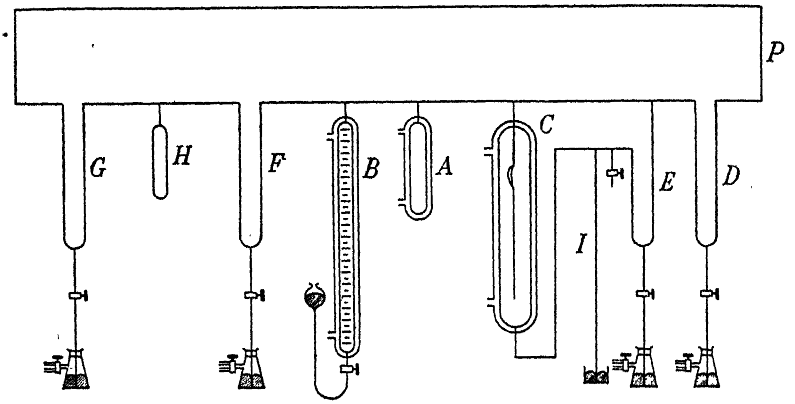


FIGURE 7. Apparatus for vapour-phase experiments.

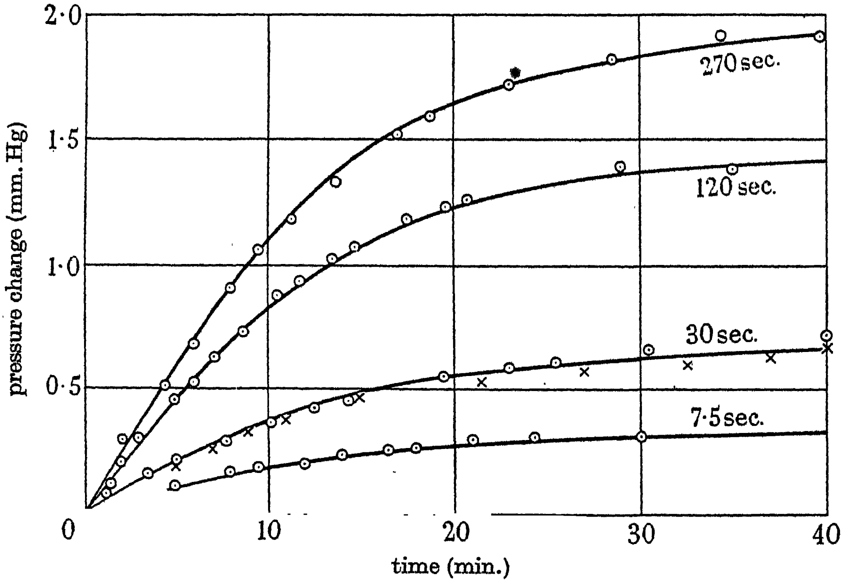


FIGURE 8. Heterogeneous polymerization, first phase; $p = 24$ mm.

(iii) After about 10^{20} molecules of monomer had been removed from *C*, the latter was cooled in liquid air for some time. A considerable fraction of the monomer removed distilled back into *C* from *A*—up to 70 % when the vapour was almost saturated, i.e. under conditions similar to Melville's. Evidently a large part of the rate observed by Melville was due not to polymerization, but to solution of the monomer by the polymer—a result which might have been anticipated. It is clear therefore that no definite conclusions about the polymerization can be drawn from such measurements.

---

* This variation is apart from the discrepancy of a power of 10 between the rates given in his table II and the experimental results in the same table.

We therefore decided to follow the early stages of the reaction in a more precise manner. The apparatus shown in figure 7 was constructed. A quartz reaction vessel A was connected to a burette B, and a sensitive glass bourdon C. All three were kept at 25° C thermostatically. Mercury cut-offs D, E, F, G were used instead of taps. Pure monomer stored in the tube H could be admitted through F to A, B, C, at a constant pressure (24 mm.) read on manometer I, using C as a null instrument. The vapour in A was irradiated by 2537 Å at constant intensity for a known time, and the reaction then followed by the pressure change read on C. The pressure changes measured were always small compared to the total pressure, and periodically the pressure was brought back to its original value by raising the mercury in B.

Three phases of the reaction were observed.

(i) In the early stages the rate of pressure change falls quite rapidly after the light has been cut off, following a first-order law, with a half-life of about 10 min., independent of the time of irradiation τ. Further, the initial rate is approximately proportional to $\tau^{\frac{1}{2}}$. This is illustrated in figure 8, in which the curves represent first-order plots with a half-life of 10 min. The experimental points refer to values of $\tau = 7{\cdot}5$, 30, 120, 270 sec. respectively. The initial rates are in the ratio 1·0 : 2·0 : 4·4 : 5·9. In this phase the results are fairly reproducible; the experiments represented in figure 8 were carried out in the order 30, 120, 270, 30, 7·5. If Melville's mechanism were correct the initial rates should be directly proportional to τ. If, however, the polymerization is of the radical type, initiated by catalyst produced photochemically, it is easily seen that the initial rate should be proportional to $\tau^{\frac{1}{2}}$. The concentration of catalyst c at constant monomer concentration will be given by

$$c = c_0\,\mathrm{e}^{-k_c t}, \tag{1}$$

and also*

$$c_0 = a\tau, \tag{2}$$

where a is constant for constant light intensity. Thus

$$-\frac{d[M]}{dt} = k_2[M]\left(\frac{B}{k_4}\right)^{\frac{1}{2}}, \tag{3}$$

where B is the rate of initiation and is equal to $k_c c$. Hence, from (1) and (2),

$$-\frac{d[M]}{dt} = k_2[M]\left(\frac{k_c c}{k_4}\right)^{\frac{1}{2}} = \text{constant}\,\tau^{\frac{1}{2}}\,\mathrm{e}^{-\frac{1}{2}k_c t}. \tag{4}$$

From (4) it also follows that the rate should obey a first-order law, with a half-life independent of τ.

(ii) When about 2×10^{18} molecules/cm.^{2} of monomer had been deposited as polymer on the walls of A, the rates became very much higher but still decreased with time, the half-life being unchanged.

---

* Strictly, allowance should be made for decay of catalyst during irradiation. However, this correction is not large enough to affect the results significantly.

(iii) With further increase in the amount of polymer in *A* (more than about 5×10^{18} molecules monomer/cm.^{2}), the initial rates decrease (as also observed by Melville), but the decrease in rate after cutting off the light becomes progressively less rapid.

Figure 9 illustrates these points. It shows pressure-time curves for the dark reaction following various periods of illumination for different amounts of polymer on the walls of *A*. These amounts are indicated in the diagram. The last curve of the series, which refers to an initial polymer deposit of 8×10^{18} molecules monomer/cm.^{2}, is virtually linear for the first 15 min., but at 24 min. there is a definite falling off in rate.

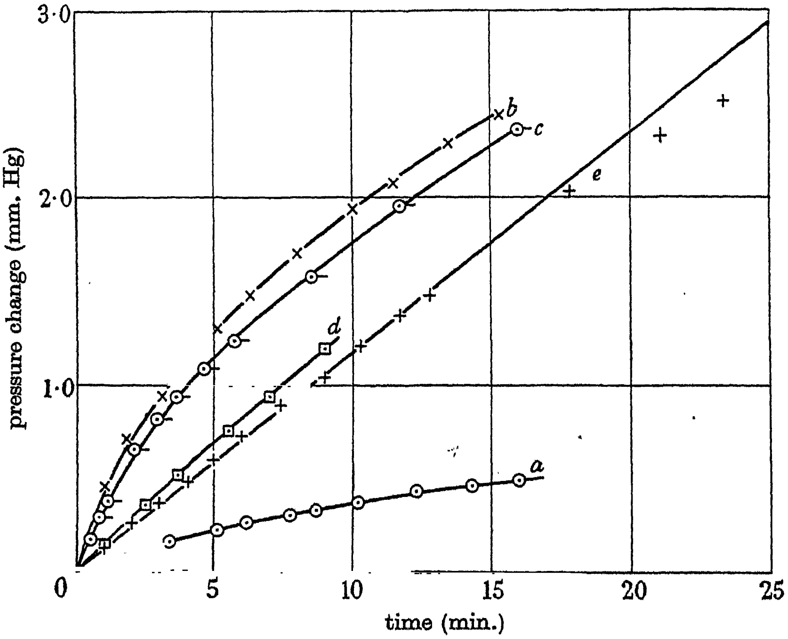


FIGURE 9. Heterogeneous polymerization, later phases; $p = 24$ mm.

curve	irradiation (sec.)	molecules of polymer/cm.^{2} initially ($\times 10^{-18}$)
a	30	0·16
b	30	3·0
c	30	5·5
d	120	7·0
e	150	8·0

These results confirm the earlier ones, and show that in the later phases of the reaction, the rate never reaches a constant value as claimed by Melville (1937).

The first phase is readily accounted for by the production of catalyst during irradiation (see (i) above). During this phase there could not be enough monomer on the walls to absorb a significant fraction of the incident light, and catalyst is therefore formed in the gas phase. If this were not so the rates corresponding to a given time of irradiation would not be independent of the amount of polymer on the walls. When phase (ii) is reached, there is sufficient monomer adsorbed on the walls to absorb most of the light. For a given intensity, catalyst may be produced more rapidly in the liquid than in the gas phase. Since the rate of reaction is still governed by the catalyst concentration, the half-life period is unchanged. Finally, when there is a large amount of polymer, the catalyst will be formed in a layer next to the wall, and diffusion of monomer through the upper layers of polymer may become rate determining. The rate would then remain sensibly constant until a large proportion of the catalyst had disappeared.

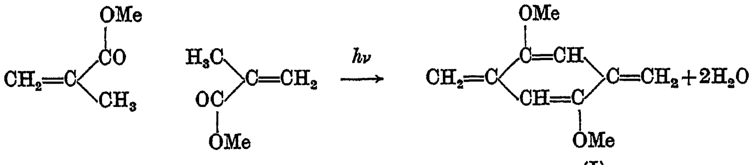
In the above experiments no correction has been applied for the absorption of monomer by the polymer. The absorption was found to be small (less than 10 %), as would be expected, since we worked with pressures of methacrylate well below the saturation pressure. Moreover, if during stages (i) and (ii) above, the monomer was removed at the end of a run, and readmitted, equilibrium was established very rapidly. Therefore during such runs the polymer was always saturated with monomer, and the observed rates were then proportional to the true rates of polymerization.

Under the conditions of Melville's experiments, there is therefore no reason to believe that the general mechanism of polymerization differs at all from the usually accepted scheme for radical chains; nor are there any grounds for believing that the polymer can have special properties, e.g. abnormally high molecular weight. In our opinion, the available evidence strongly supports the photochemical formation of a catalyst which we have postulated.

It seems very likely that a similar explanation will be found to hold for the vapour-phase polymerization of chloroprene. Chloroprene, according to Bolland & Melville (1938, 1939), shows a similar photochemical after-effect, which they ascribe to the 'double-bond mechanism'.

NATURE OF THE PHOTOCATALYST AND THERMAL INHIBITOR

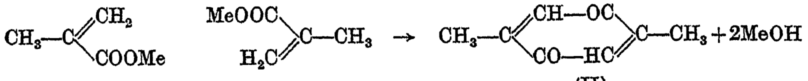
The formation of catalyst and inhibitor from pure methyl methacrylate has been postulated on kinetic grounds, and it would be desirable to have direct chemical evidence for the existence of these substances. We have not been able to identify them, but it is not chemically unreasonable that compounds with such properties should be formed. Thus the catalyst might have the structure (I), being formed by the reaction



(I)

This reaction is similar to the photocondensation of acetone and cyclohexane (Bowen & Horton 1936). The work of Szwarc (1947) suggests that (I) would be an effective catalyst. (This possibility is being investigated.) Incidentally, similar reactions can be written for chloroprene and methyl isopropenyl ketone, which are at present the only other compounds known to show the abnormal photochemical after-effect (Melville 1941). In these cases the catalysts would be (I) with methoxyl replaced by hydrogen or methyl respectively.

The inhibitor could be *p*-xyloquinone (II), formed by the reaction



(II)

The work of Walling & Briggs (1946) suggests that the compound is a strong inhibitor but a weak retarder, and (II) should show such behaviour.

DETERMINATION OF THE VELOCITY CONSTANTS

The method described in part I can be used to determine the velocity constants provided suitable precautions are taken. First, the measurements must be made at low temperatures, so that the formation of inhibitor during the experiments is not significant. Our measurements were carried out at 0° C. Secondly, the variation in rate due to decay of the photochemical catalyst must be small over the period required for measuring the after-effect. Thirdly, since irradiation produces a catalyst, it is impossible to measure the after-effect for a final rate corresponding to the true thermal rate. However, it is possible to adjust the initial concentration of catalyst so that the rate after measurement of the after-effect has a constant value.* This value, the 'effective thermal rate', is used in place of the thermal rate in the equations to determine k_2, k_3 and k_4; k_1 is then found by substituting these values in the expression for the true thermal rate. At 0° C the thermal rate has the value $2{\cdot}6 \times 10^{-5}$ sec.$^{-1}$ (see figure 4), and the effective thermal rate used was $3{\cdot}3 \times 10^{-5}$ sec.$^{-1}$. The time required to measure the after-effect was 15 min. In this time the effective thermal rate falls to $3{\cdot}14 \times 10^{-5}$ sec.$^{-1}$. The error so introduced into the after-effect is less than 5 %.

KINETICS AND DETERMINATION OF THE VISCOSITY EXPONENT

The kinetic scheme developed in part I leads to the following expressions:

$$\bar{P}_n = \frac{2k_2(k_1k_4)^{-\frac{1}{2}}}{2\beta+\xi}, \tag{5}$$

$$\frac{d\eta}{dt} = KM_0^{\alpha}[M]^2\left(\frac{k_2^2}{k_4}\right)^{\frac{1}{2}(1+\alpha)} k_1^{\frac{1}{2}(1-\alpha)} \frac{\xi[\beta+(1+\frac{1}{2}\alpha)\,\xi]}{(\beta+\xi)^{1+\alpha}}, \tag{6}$$

where $\beta = k_3(k_1k_4)^{-\frac{1}{2}}$, $\xi = [1+AI/k_1[M]^2]^{\frac{1}{2}}$, M_0 is the molecular weight of the monomer, $\bar{P}$ is the number average degree of polymerization, I is the light intensity in arbitrary units, A is constant for a given temperature and experimental set up, η is the ideal specific viscosity, K, α are the constants relating intrinsic viscosity to molecular weight in the Houwink equation

$$[\eta] = KM^{\alpha}, \tag{7}$$

---

* This could only be done by trial and error, but the new design of viscometer made it possible to carry out the very large number of measurements required with a single specimen of monomer.

and k_1, k_2, k_3, k_4 are the velocity constants of initiation, propagation, chain transfer and termination, respectively.

In order to proceed further it is necessary to evaluate K and α. The latter may be estimated by measuring the relation between $d\eta/dt$ and I for very high light intensities. Under these conditions equation (6) reduces to

$$\frac{d\eta}{dt} = \text{constant}\, I^{\frac{1}{2}(1-\alpha)}. \tag{8}$$

This constitutes a new method for determining α.

In figure 10, $\log I$ is plotted against $\log d\eta/dt$. The last five points lie satisfactorily on a straight line of slope 0·115 giving $\alpha = 0{\cdot}77$. As shown in part I, the kinetic equations are relatively insensitive to α; since the expression for the after-effect can only be obtained explicitly for rational α, we have taken $\alpha = \frac{4}{5}$. This seemed preferable to $\frac{3}{4}$, since the method just described may give values for α which are slightly too small. Moreover, the choice $\alpha = 0{\cdot}8$ is entirely justified by the data in table 1 below.

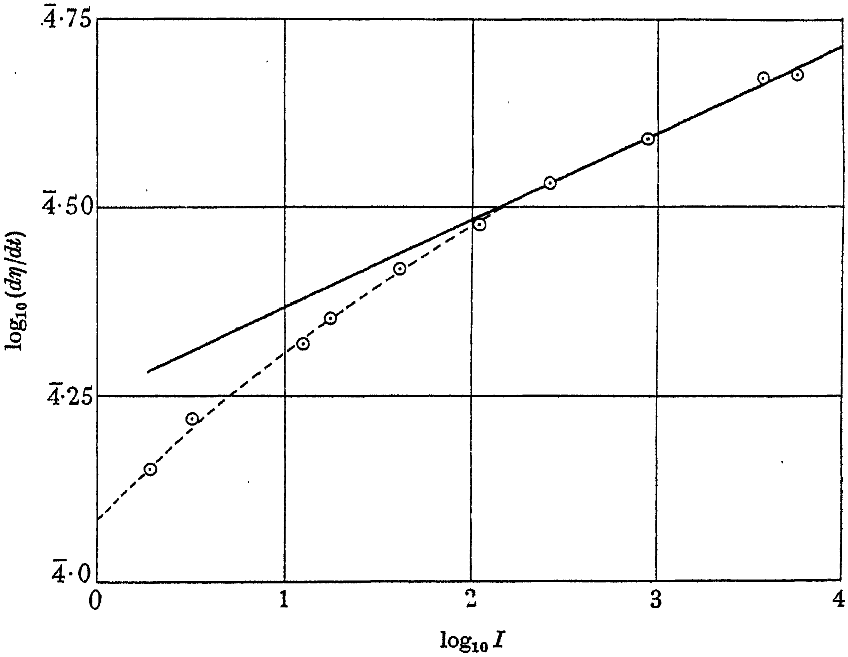


FIGURE 10. Determination of α; theoretical intensity-rate curve and experimental points.

Baxendale, Bywater & Evans (1946) obtained values of K and α by comparison of osmotic and viscosity data on fractionated polymethyl methacrylate. They found

$$\left.\begin{array}{ll} \text{in benzene} & \bar{P} = 2{\cdot}81 \times 10^3[\eta_g]^{1{\cdot}32}, \\ \text{in chloroform} & \bar{P} = 1{\cdot}81 \times 10^3[\eta_g]^{1{\cdot}22}. \end{array}\right\} \tag{9}$$

$[\eta_g]$ is the intrinsic viscosity, with concentrations expressed in g./100 cm.^{3}. From the relation established in part II (1948) between the constants K for homogeneous and unfractionated polymer, it follows that for unfractionated polymer

$$\left.\begin{aligned} &\text{in benzene} && [\eta_g] = 3{\cdot}94\times 10^{-3}\bar{P}_n^{0{\cdot}758},\\ &\text{in chloroform} && [\eta_g] = 3{\cdot}63\times 10^{-3}\bar{P}_n^{0{\cdot}820}.\end{aligned}\right\} \tag{10}$$

A photochemical polymer was prepared at 0° C and values of $[\eta_g]$ were determined in benzene and in methyl methacrylate. These were 10·04 and 7·925 respectively. Hence from (10), $\bar{P}_n = 3{\cdot}125\times 10^4$, and thus, in monomer solution, assuming $\alpha = 0{\cdot}8$ and with concentration expressed in base moles/litre (the unit used throughout this paper)

$$K = 5{\cdot}05\times 10^{-4}. \tag{11}$$

By introducing $\alpha = \frac{4}{5}$ into the general equations given in part I for the after- and pre-effects, and integrating, we find

$$\begin{aligned}\Delta\eta_I^\infty = -\frac{KM_0^{\frac{4}{5}}[M]\,k_2^{\frac{9}{5}}}{2k_1^{\frac{2}{5}}k_4^{\frac{7}{5}}}\Bigg\{&\frac{\beta+1{\cdot}4}{\lambda^9}\Bigg[1{\cdot}863\log_{10}\frac{\psi^2+1{\cdot}618\psi\lambda+\lambda^2}{3{\cdot}618\lambda^2}\\ &+1{\cdot}175\tan^{-1}\frac{2\psi+1{\cdot}618\lambda}{1{\cdot}175\lambda}-0{\cdot}7116\log_{10}\frac{\psi^2-0{\cdot}618\psi\lambda+\lambda^2}{1{\cdot}832\lambda^2}\\ &-1{\cdot}902\tan^{-1}\frac{2\psi-0{\cdot}618\lambda}{1{\cdot}902\lambda}-2{\cdot}303\log_{10}\frac{2{\cdot}5\lambda^4(\psi-1)(\xi+1)}{\xi-1}-0{\cdot}367\Bigg]\\ &+\frac{\beta-1{\cdot}4}{\nu^9}\Bigg[1{\cdot}863\log_{10}\frac{\psi^2+1{\cdot}618\psi\nu+\nu^2}{\lambda^2+1{\cdot}618\lambda\nu+\nu^2}+1{\cdot}175\tan^{-1}\frac{2\psi+1{\cdot}618\nu}{1{\cdot}175\nu}\\ &-0{\cdot}7116\log_{10}\frac{\psi^2-0{\cdot}618\psi\nu+\nu^2}{\lambda^2-0{\cdot}618\lambda\nu+\nu^2}-1{\cdot}902\tan^{-1}\frac{2\psi-0{\cdot}618\nu}{1{\cdot}902\nu}\\ &-2{\cdot}303\log_{10}\frac{\psi-\nu}{\lambda-\nu}-0{\cdot}150\Bigg]-\frac{\beta^2}{\beta^2-1}\left[\frac{1}{\lambda^4}-\frac{1}{\psi^4}\right]\Bigg\},\end{aligned} \tag{12}$$

where $\Delta\eta_I^\infty$ is the limiting after-effect for intensity I and

$$\lambda^5 = \beta+1,\quad \nu^5 = \beta-1,\quad \psi^5 = \beta+\xi.$$

Also

$$\begin{aligned}\Delta'\eta_I^\infty = -\frac{KM_0^{\frac{4}{5}}[M]\,k_2^{\frac{9}{5}}}{2k_1^{\frac{2}{5}}k_4^{\frac{7}{5}}}\Bigg\{&\frac{\beta+1{\cdot}4\xi}{\psi^9}\Bigg[1{\cdot}863\log_{10}\frac{\psi^2+1{\cdot}618\psi\lambda+\lambda^2}{3{\cdot}618\psi^2}\\ &+1{\cdot}175\tan^{-1}\frac{2\lambda+1{\cdot}618\psi}{1{\cdot}175\psi}-0{\cdot}7116\log_{10}\frac{\psi^2-0{\cdot}618\psi\lambda+\lambda^2}{1{\cdot}382\psi^2}\\ &-1{\cdot}902\tan^{-1}\frac{2\lambda-0{\cdot}618\psi}{1{\cdot}902\psi}-2{\cdot}303\log_{10}\frac{2{\cdot}5\psi^4(\psi-\lambda)(\xi+1)}{\xi(\xi-1)}-0{\cdot}367\Bigg]\\ &+\frac{\beta-1{\cdot}4\xi}{\chi^9}\Bigg[1{\cdot}863\log_{10}\frac{\lambda^2+1{\cdot}618\lambda\chi+\chi^2}{\psi^2+1{\cdot}618\psi\chi+\chi^2}+1{\cdot}175\tan^{-1}\frac{2\lambda+1{\cdot}618\chi}{1{\cdot}175\chi}\\ &-0{\cdot}7116\log_{10}\frac{\lambda^2-0{\cdot}618\lambda\chi+\chi^2}{\psi^2-0{\cdot}618\psi\chi+\chi^2}-1{\cdot}902\tan^{-1}\frac{2\lambda-0{\cdot}618\chi}{1{\cdot}902\chi}\\ &-2{\cdot}303\log_{10}\frac{\lambda-\chi}{\psi-\chi}-0{\cdot}150\Bigg]-\frac{\beta^2}{\xi^2-\beta^2}\left[\frac{1}{\lambda^4}-\frac{1}{\psi^4}\right]\Bigg\},\end{aligned} \tag{13}$$

where $\Delta'\eta_I^\infty$ is the pre-effect for intensity I, and $\chi^5 = \beta-\xi$.

RESULTS

A thermal polymer at 0° C was prepared by irradiating the monomer in a viscometer, and allowing to stand until the rate had fallen to the thermal value (see figure 4). The liquid was then poured into the appendage, and the pure monomer distilled back and allowed to polymerize at 0° C. The polymer was isolated by precipitation with methanol, and its intrinsic viscosity determined in benzene. The value found was 163. Thus, at 0° C

$$\bar{P}_n = 5{\cdot}93 \times 10^4. \tag{14}$$

The method used for determining photochemical rates and after-effects was similar to that described in part I, with the modification mentioned earlier. The photochemical rates are given in table 1, as ratios to the effective thermal rate.

TABLE 1

	ratio to effective thermal rate	
I	obs.	calc.
0·097	1·84	1·81
0·613	3·21	3·21
1·87	4·32	4·38
3·20	5·00	4·98
12·40	6·35	6·56
17·50	6·54	6·96
40·70	7·90	7·95
105·0	9·15	9·06
255·0	10·32	10·12
891·0	11·86	11·86
3680·0	14·30	13·60
5540·0	14·49	14·20

The calculated values given in table 1 were obtained with

$$A = 32{\cdot}1, \quad \beta' = 3{\cdot}5, \tag{15}$$

($\beta' = k_3(k_1' k_4)^{-\frac{1}{2}}$, where k_1' refers to the effective thermal rate). The agreement is extremely good over a range of light intensities of 50,000 : 1, and rates of chain starting of 180,000 : 1. This may be seen clearly from figure 10; the dotted line is the theoretical curve.

From equations (6) and (11), with $[M] = 9{\cdot}62$, it follows that

$$(k_2^2/k_4)^{0{\cdot}9}\,(k_1')^{0{\cdot}1} = 5{\cdot}40 \times 10^{-5}. \tag{16}$$

The observed values of $\Delta\eta_I^\infty - \Delta'\eta_I^\infty = \delta\eta_I$ are given in table 2. The calculated values of $\delta\eta_I$ and $\Delta'\eta_I^\infty$ were obtained from equations (12) and (13) with $[M] = 9{\cdot}62$ and

$$\frac{KM_0^{\frac{4}{5}}k_2^{\frac{8}{5}}}{2(k_1')^{\frac{3}{5}}k_4^{\frac{7}{5}}} = 3{\cdot}10 \times 10^{-3}. \tag{17}$$

The values of $\delta\eta$ at high light intensities are the more reliable.

From equations (5), (14), (15), (16), (17) and substituting the true thermal rate ($2{\cdot}6 \times 10^{-5}$ sec.$^{-1}$) in (6) we find the values of the velocity constants k_1, k_2, k_3, k_4 and k_1' given in table 3.

Table 4 shows calculated values for the lifetime of single radicals, of growing polymer molecules, and of kinetic chains.

Table 2

I	$\delta\eta_I$ (obs. × 10^2)	$\delta\eta_I$ (calc. × 10^2)	$\Delta'\eta_I^\infty$ (calc. × 10^2)
0·613	1·41	1·78	0·51
3·20	1·74	2·07	0·36
17·5	2·43	2·58	0·19
105·0	3·03	3·03	0·09
891·0	3·22	3·20	0·06
5540·0	3·44	3·38	0·02

Table 3

	(l.mol.$^{-1}$sec.$^{-1}$)
k_1	$6{\cdot}82\times10^{-15}$
k_2	41·6
k_3	$6{\cdot}13\times10^{-4}$
k_4	$2{\cdot}69\times10^{6}$
k_1'	$1{\cdot}14\times10^{-14}$
β	4·51

Table 4

mean lifetimes at 0° C in thermal polymerization

single radical	polymer molecule	kinetic chain
$2{\cdot}5\times10^{-3}$ sec.	148 sec.	816 sec.

Discussion

The only quantitative results in the liquid phase available for comparison are those of Schulz & Blaschke (1942), Schulz & Harborth (1947) and Baxendale, Bywater & Evans (1946). Schulz *et al.* measured overall rates and molecular weights for the peroxide-catalyzed reaction in benzene solution. Under these conditions polymerization occurs through single-headed radicals, and the degree of polymerization is given by the simple expression

$$\bar{P}_n = \frac{k_2[M]}{k_3[M]+(Bk_4)^{\frac{1}{2}}}, \tag{18}$$

where B is the rate of chain starting, and is proportional to the benzoyl peroxide concentration for a constant monomer concentration. The overall rate is given by

$$-\frac{d[M]}{dt} = k_2[M]\left(\frac{B}{k_4}\right)^{\frac{1}{2}}. \tag{19}$$

By measuring $\bar{P}$ and $-d[M]/dt$ at various peroxide concentrations, and eliminating B from equations (18), (19), one can evaluate k_2^2/k_4 and k_3/k_2. Figure 11 shows Arrhenius plots for these quantities, including Schulz & Harborth's results for 50 and 70° C and our own for 0° C. The agreement is quite satisfactory in both cases, and the corresponding activation energy differences are

$$2E_2 - E_4 = 9{\cdot}9\,\text{kcal.},\quad E_3 - E_2 = 5{\cdot}9\,\text{kcal.} \tag{20}$$

The earlier and less accurate measurements of Schulz & Blaschke (1942) lead to values of 9·4 and 4·1 kcal. respectively.

The corresponding figures for styrene are 10·2 and 7·7 kcal. (part I). It seems likely that E_2 and E_4 are similar for both reactions, since co-polymerization experiments have shown that there is no great difference in the reactivities of the radicals or monomers. Probably E_3 is somewhat smaller for methyl methacrylate.

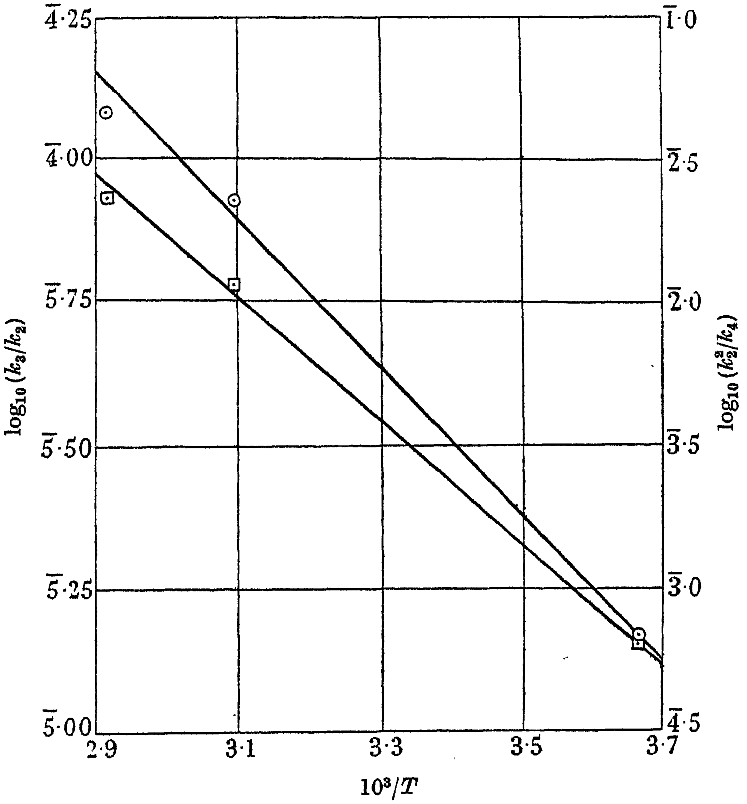


FIGURE 11. Arrhenius plots for methyl methacrylate. ⊙ $\log_{10}(k_3/k_2)$, ⊡ $\log_{10}(k_2^2/k_4)$.

Baxendale, Evans & Kilham (1946) estimated $k_2/k_4^{\frac{1}{2}}$ for the polymerization in aqueous solution, initiated by hydroxyl radicals, making use of equation (19) and substituting for B the rate of production of hydroxyl radicals, which they measured independently. They found for $k_2/k_4^{\frac{1}{2}}$ the value 12·3 at 25°, which differs from that estimated from figure 11 ($5{\cdot}2\times10^{-2}$) by a factor of about 240. This discrepancy could be ascribed to the Norrish & Smith (1942) effect; in a poor solvent the growing polymer chains coil up and the termination constant is reduced. The effect should be particularly large in water. Moreover, the failure of the line in their figure 3A to pass through the origin can be so explained. Judging by Norrish & Smith's and Trommsdorff's (1944) results, the decrease in k_4 should set in only when the concentration of polymer exceeds a certain value depending on the solvent. The line plotted by Baxendale *et al.* should bend up sharply at this point, and so the final linear position will not pass through the origin. Baxendale *et al.* record no measurements of rate during this preliminary phase. From Norrish & Smith's results it can be calculated that if the increase in rate in bad solvents is due entirely to a decrease in k_4, then k_4 decreases by a factor of at least 100 in amyl acetate, and by a much larger factor in cyclohexanol. The variation observed in water is therefore reasonable.

Melville (1937) tried to measure the lifetime of growing chains in the hydrogen-sensitized vapour-phase polymerization by the rotating sector method of Briers, Chapman & Walters (1926). If the active centres under his conditions were mostly growing polymer radicals, his data would lead to a value of 3×10^6 l.mol.$^{-1}$sec.$^{-1}$ for k_2. This value is improbably high, and it therefore seems likely that the majority of active centres must have been hydrogen atoms, and not polymer radicals.

THE TERMINATION REACTION

Both combination and disproportionation of radicals have been suggested as possible mechanisms for termination, and doubt is still expressed in some quarters on this point. Our results seem to lead to a definite decision in the cases of styrene and methyl methacrylate. The following simple treatment is sufficient to show that the observations are incompatible with a combination mechanism. The more detailed analysis is not reproduced, as we do not think that this case will ever arise in practice.

Using the kinetic scheme of part I, with appropriate modifications for the termination reactions, we may show that the total concentration of radicals $[X]_I$ is given by

$$[X]_I = \xi[M]\,(k_1/k_4)^{\frac{1}{2}}, \tag{21}$$

and hence the rate of disappearance of monomer by

$$-\frac{d[M]}{dt} = k_1\xi^2[M]^2 + (k_2+k_3)\,[M]\,[X] \sim k_2\xi[M]^2\left(\frac{k_1}{k_4}\right)^{\frac{1}{2}}. \tag{22}$$

These expressions are identical with those holding for termination by disproportionation. The ends of the finished molecules can only be formed by the transfer reaction, each transfer process producing two ends. Since each molecule has two ends, the rate of production of polymer molecules is

$$\frac{d[P]}{dt} = k_3\xi[M]^2\left(\frac{k_1}{k_4}\right)^{\frac{1}{2}}. \tag{23}$$

From (22) and (23) we obtain for the number average degree of polymerization

$$\overline{P}_n = k_2/k_3. \tag{24}$$

The rate of viscosity increase during illumination is then given by

$$\frac{d\eta}{dt} = KM_0[M]^2\,\xi\left(\frac{k_1}{k_4}\right)^{\frac{1}{2}} k_2^{1+\alpha}k_3^{-\alpha}, \tag{25}$$

provided the viscosity average is proportional to the number average. The detailed treatment, following the method of part II, shows that

$$1 < \overline{P}_{\text{visc.}}/\overline{P}_n \leqslant 2 \tag{26}$$

for $0 < \alpha \leqslant 1$.

Hence if termination occurred by combination, the molecular weight would be independent of the rate of chain starting, and the rate of viscosity increase in the photopolymerization would be proportional to $I^{\frac{1}{2}}$ at high light intensities. We find, however, that the molecular weight decreases with increasing rate of initiation (see part I and pp. 368–9 *supra*), and that at the highest intensities used, $d\eta/dt$ has the values 0·23 (styrene at 0° C) and 0·115 (methyl methacrylate at 0°). This discrepancy is much too large to be attributed to variations in the ratio $\bar{P}_{\text{visc.}} : \bar{P}_n$, as will be obvious from equation (26) and figure 10. It will be clear from the present results, and from those described in part I, that both methyl methacrylate and styrene follow closely the kinetics calculated on the basis of termination by disproportionation (see especially table 1 in this paper, in which the rate of initiation varies by a factor of nearly 2×10^5).

It therefore seems certain that termination occurs by disproportionation in the polymerization of styrene and methyl methacrylate under our conditions; and similar conclusions have been reached by Dr Axford and Dr Dixon Lewis working in these laboratories on the polymerizations of *p*-methoxystyrene and of vinyl acetate.

Baxendale, Bywater & Evans (private communication and 1946) have given evidence which appears to show that during the polymerization of methyl methacrylate in aqueous solution initiated by Fenton's reagent termination occurs by combination of radicals. This may imply that in aqueous solution the termination mechanism is qualitatively different from that in pure monomer, but since the conditions used by Baxendale, Bywater & Evans were extremely complex, other explanations of their experiments could be suggested.

REFERENCES

Bamford, C. H. & Dewar, M. J. S. Part I 1948*a* *Proc. Roy. Soc.* A, **192**, 309.
Bamford, C. H. & Dewar, M. J. S. Part II 1948*b* *Proc. Roy. Soc.* A, **192**, 329.
Baxendale, J. H., Bywater, S. & Evans, M. G. 1946 *J. Polymer Sci.* **1**, 237.
Baxendale, J. H., Evans, M. G. & Kilham, J. K. 1946 *Trans. Faraday Soc.* **42**, 668.
Bolland, J. L. & Melville, H. W. 1938 *Proc. Rubber Tech. Conf. London*, p. 239.
Bolland, J. L. & Melville, H. W. 1939 *Öst. ChemZtg*, no. 10.
Bowen, E. J. & Horton, A. T. 1936 *J. Chem. Soc.* p. 1685.
Briers, F., Chapman, D. L. & Walters, E. 1926 *J. Chem. Soc.* p. 562.
Foord, S. G. 1940 *J. Chem. Soc.* p. 48.
Melville, H. W. 1937 *Proc. Roy. Soc.* A, **163**, 511.
Melville, H. W. 1939 *I.R.I. Trans.* **15**, 209.
Melville, H. W. 1940 *Ann. Rep. Chem. Soc.* **37**, 68.
Melville, H. W. 1941 *J. Chem. Soc.* p. 414.
Norrish, R. G. W. & Brookman, E. F. 1939 *Proc. Roy. Soc.* A, **171**, 147.
Norrish, R. G. W. & Smith, R. R. 1942 *Nature*, **150**, 336.
Schulz, G. V. & Blaschke, F. 1942 *Z. phys. Chem.* **51** B, 75.
Schulz, G. V. & Harborth, G. 1947 *Makromol. Chem.* **1**, 106.
Schulz, G. V. & Sing, G. 1943 *J. prakt. Chem.* **161**, 161.
Szwarc, M. 1947 *Disc. Faraday Soc.* no. 2, 46.
Trommsdorff, E. 1944 *Colloquium on high polymers*, Freiburg. See *B.I.O.S. Report*, no. 363, item 22.
Walling, C. & Briggs, E. R. 1946 *J. Amer. Chem. Soc.* **68**, 1141.

Studies in polymerization

IV. The velocity constants in the polymerization of *p*-methoxystyrene

By D. W. E. Axford

Courtaulds Ltd. Research Laboratory, Maidenhead, Berks

(*Communicated by A. H. Wilson, F.R.S.—Received* 14 *December* 1948)

The velocity constants in the polymerization of *p*-methoxystyrene have been determined at 0° C by the method described in part I. The activation energies could not be measured directly. Comparison of the results with those for styrene suggests that the differences are due to mesomeric stabilization of the radicals by the *p*-substituent.

The relation between intrinsic viscosity and molecular weight of the polymer has been studied by the method of part II. The results conform to the Houwink (1940) viscosity law, with an exponent 0·66, almost identical with that for polystyrene. If the Houwink equation is written in the form $[\eta] = C\bar{P}^{\alpha}$, C is the same for solutions of both polymers in their respective monomers.

Introduction

In parts I and II of this series, Bamford & Dewar (1948*a*, *b*) have described a method of determining the absolute velocity constants in the liquid-phase polymerization of vinyl compounds. The method is now being applied in these laboratories to a series of monomers, with the object of elucidating the relation between reactivity and structure. The present paper described such an investigation of *p*-methoxystyrene, a substance which resembles styrene in its general behaviour, and for which the techniques described in parts I and II could be taken over without modification.

Summary of the method

For clarity, a brief summary of the method is given. In order to determine the four velocity constants four quantities must be measured. These are (i) the rate of viscosity change when the liquid monomer is irradiated with light of various intensities (this provides two relations between the velocity constants, and also the proportionality factor between the incident light intensity and the rate of initiation); (ii) the degree of polymerization of the thermal polymer; (iii) the photochemical after-effect.

The kinetic scheme is as follows:

$$\begin{array}{lll}
2M \to 2X \text{ or } D_2 & (k_1) & \text{initiation,}\\
X+M \to X & (k_2) & \text{propagation,}
\end{array}$$

$$\left.\begin{array}{ll}
D_2+M \to D_1+R & (2k_3)\\
D_1+M \to P_2+R & (k_3)\\
R+M \to P_1+R & (k_3)
\end{array}\right\}\ \text{transfer,}$$

$$\left.\begin{array}{ll}
D_2+X \to D_1+Q & (2k_4)\\
D_1+X \to P_2+Q & (k_4)\\
R+X \to P_1+Q & (k_4)\\
2X \to 2Q & (k_4)
\end{array}\right\}\ \text{termination.}$$

X represents any active centre and Q any dead centre. D_2 is the initial polymer growing at both ends, D_1 that growing at one end only, and R the growing transfer polymer. P_1 and P_2 are dead transfer and initial polymers respectively.

(i) To calculate the rate of change of viscosity on the basis of this scheme, it is necessary to know the relation between viscosity, concentration and mean degree of polymerization. In part II of this series (Bamford & Dewar 1948*b*), it was shown that for a heterogeneous polymer formed by the growth of a single radical centre (a transfer or catalytic polymer), the equation

$$[\eta] = K'(M_0\bar{P})^{\alpha} \tag{1}$$

holds, where M_0 is the molecular weight of the monomer, and $\bar{P}$ is the number average degree of polymerization. For a polymer formed by the growth of a double radical (a thermal or photochemical polymer),

$$[\eta] = K''(M_0\bar{P})^{\alpha}, \tag{2}$$

where $K'' = K'(2+\alpha)/2^{1+\alpha}$. In the next section it will be shown that for poly-*p*-methoxystyrene, $K' = 3{\cdot}68\times10^{-3}$, $\alpha = 0{\cdot}66$. These relationships, together with the reaction scheme given above, lead to the following expression for the rate of change of viscosity* in a photochemical polymerization:

$$\frac{d\eta}{dt} = K'M_0^{\alpha}[M]^2\left(\frac{k_2^2}{k_4}\right)^{\frac{1}{2}(1+\alpha)} k_1^{\frac{1}{2}(1-\alpha)}\frac{\xi\{\beta+\xi(1+\frac{1}{2}\alpha)\}}{(\beta+\xi)^{1+\alpha}}, \tag{3}$$

where $[M]$ is the monomer concentration, $\xi = (1+AI)^{\frac{1}{2}}$, I being the light intensity in arbitrary units, A is a constant, and $\beta = k_3/(k_1k_4)^{\frac{1}{2}}$. The rate of change of viscosity in a thermal polymerization can be obtained from this by putting $I = 0$. Measurement of $d\eta/dt$ over a range of light intensities enables β, $(k_2^2/k_4)^{\frac{1}{2}(1+\alpha)}\,k_1^{\frac{1}{2}(1-\alpha)}$, and A to be determined.

(ii) The number average degree of polymerization of the thermal polymer is given by

$$\bar{P} = \frac{2k_2(k_1k_4)^{-\frac{1}{2}}}{2\beta+1}. \tag{4}$$

(iii) The expression for the photochemical after- and pre-effects are as follows (Bamford & Dewar, 1948*a*), taking $\alpha = 2/3$ (see §(*c*)):

The total after-effect:

$$\begin{aligned}\Delta\eta_I^{\infty} = -\tfrac{1}{2}K'M_0^{\frac{2}{3}}[M]k_1^{-\frac{1}{3}}k_2^{\frac{5}{3}}k_4^{-\frac{4}{3}}\Bigg[&\frac{\beta+\frac{4}{3}}{\lambda^5}\left\{\tfrac{1}{2}\log\frac{4(\xi-1)^3}{27\lambda^6(\xi+1)^2(\psi-\lambda)^3}\right.\\ &\left.-\sqrt{3}\tan^{-1}\sqrt{3}+\sqrt{3}\tan^{-1}\frac{2\psi+\lambda}{\sqrt{3}\,\lambda}\right\}+\frac{\beta-\frac{4}{3}}{\nu^5}\left\{\tfrac{1}{2}\log\frac{(\lambda-\nu)^3(\xi+1)}{2(\psi-\nu)^3}\right.\\ &\left.-\sqrt{3}\tan^{-1}\frac{2\lambda+\nu}{\sqrt{3}\nu}+\sqrt{3}\tan^{-1}\frac{2\psi+\nu}{\sqrt{3}\nu}\right\}-\frac{\beta^2}{\lambda^3\nu^3}\left\{\frac{1}{\lambda^2}-\frac{1}{\psi^2}\right\}\Bigg].\end{aligned} \tag{5}$$

The total pre-effect:

$$\begin{aligned}\Delta\eta_I'^{\infty} = -\tfrac{1}{2}K'M_0^{\frac{2}{3}}[M]k_1^{-\frac{1}{3}}k_2^{\frac{5}{3}}k_4^{-\frac{4}{3}}\Bigg[&\frac{\beta+\frac{4}{3}\xi}{\chi^5}\left\{\tfrac{1}{2}\log\frac{4\xi^2(\xi-1)^3}{27\psi^6(\xi+1)^2(\psi-\lambda)^3}\right.\\ &\left.-\sqrt{3}\tan^{-1}\sqrt{3}+\sqrt{3}\tan^{-1}\frac{2\lambda+\psi}{\sqrt{3}\,\psi}\right\}+\frac{\beta-\frac{4}{3}\xi}{\chi^5}\left\{\tfrac{1}{2}\log\frac{(\xi+1)(\psi-\chi)^3}{2\xi(\lambda-\chi)^3}\right.\\ &\left.-\sqrt{3}\tan^{-1}\frac{2\psi+\chi}{\sqrt{3}\,\chi}+\sqrt{3}\tan^{-1}\frac{2\lambda+\chi}{\sqrt{3}\,\chi}\right\}+\frac{\beta^2}{\psi^3\chi^3}\left\{\frac{1}{\lambda^2}-\frac{1}{\psi^2}\right\}\Bigg],\end{aligned} \tag{6}$$

where $\lambda = (\beta+1)^{\frac{1}{3}}$, $\nu = (\beta-1)^{\frac{1}{3}}$, $\psi = (\beta+\xi)^{\frac{1}{3}}$ and $\chi = (\beta-\xi)^{\frac{1}{3}}$.

---

* Throughout this paper η represents the ideal specific viscosity, i.e. $c[\eta]$, where c is the base molar concentration.

EXPERIMENTAL

(a) *Preparation and purification of p-methoxystyrene*

This substance is not available commercially in this country, and it was therefore necessary to synthesize it. *p*-Methoxycinnamic acid is readily obtained pure and in good yield by the condensation of anisaldehyde and malonic acid (see Johnson 1942). After careful washing and drying, the acid was decarboxylated in quinoline in the presence of finely powdered copper oxide (Walling & Wolfstirn 1947). Thorough washing with dilute hydrochloric acid and distillation under reduced pressure provided good specimens of the substituted styrene, as shown by analysis (observed C, 80·5, H, 7·5; calculated for $C_9H_{10}O$, C, 80·6, H, 7·5 %) and measurements of the refractive index, the observed value being $n^{20}_D = 1{\cdot}5604$, while the value given in the literature is $n^{20}_D = 1{\cdot}5608$ (Frank & Smith 1946). The final purification was carried out in high vacuum. The monomer was heated at 100° C for a few hours and irradiated with the full light of a mercury arc for 1 hr. After heating *in vacuo* at 60° C for another 2 hr. it was distilled on to clean sodium metal, allowed to stand for 2 hr., and then fractionated. It was found necessary to heat the line electrically to enable distillation to proceed at a reasonable rate. The viscometers were filled in an apparatus constructed entirely of quartz, without taps or greased joints, and previously degassed by heating to red-heat under vacuum.

(b) *The relationship between intrinsic viscosity and molecular weight for poly-p-methoxystyrene*

The constants K' and α in equation (1) were next determined in the same way as for polystyrene (part II, Bamford & Dewar 1948*b*). Known volumes of pure *p*-methoxystyrene and pure carbon tetrachloride were mixed in glass tubes under nitrogen. The tubes were then sealed and heated in a thermostat at 60° C until about 5 % of the monomer had polymerized. The polymers were precipitated with a large excess of methyl alcohol, washed thoroughly with more alcohol and dried. Their intrinsic viscosities were measured in toluene at 25° C, using an Ostwald no. 1 viscometer, and were extrapolated to zero concentration by the method of Schulz & Sing (1943).

Mayo (1943) showed that the degrees of polymerization of polymers prepared in solution follow the equation

$$\frac{1}{\bar{P}} = \frac{1}{\bar{P}_0} + \frac{k_3'\phi}{k_2}, \tag{7}$$

where $\bar{P}$ is the number average degree of polymerization of the solvent polymer and $\bar{P}_0$ that of polymer prepared from pure monomer at the same temperature, $\phi = [S]/[M]$, where $[S]$ is the concentration of solvent, and k_3' is the velocity constant of transfer to solvent. From equations (1) and (7) it follows that $[\eta]^{-1/\alpha}$ should be a linear function of ϕ, and in part II a technique was described for finding α from this relation. It was there shown that a plot of $\log f^{\alpha}([\eta])/K'$ against $\log\phi$, where

$$\frac{f^{\alpha}[(\eta])}{K'} = \frac{1}{[\eta]} - \frac{\alpha[\eta]^{(1-\alpha)/\alpha}}{[\eta]_0^{1/\alpha}}, \tag{8}$$

should be a straight line with gradient α. The function $f^{\alpha}([\eta])$ can be calculated sufficiently accurately from (8) using an approximate value of α, since the second term on the right is normally very small. The experimental points are plotted in figure 1 and are best fitted by a line of slope 0·66. Since each polymer molecule formed in these reactions (i.e. polymerization in carbon tetrachloride solution) contains four atoms of chlorine, analysis enables the number average molecular weight to be found. The results of the analyses are given in table 1. These results show that solutions of unbranched poly-*p*-methoxystyrene in toluene obey the Houwink law, with an exponent of 0·66. For unfractionated polymers formed by the growth of single radicals, $K' = 3{\cdot}17 \times 10^{-3}$. A comparison of the intrinsic viscosities of a given polymer in toluene and in pure monomer showed that, if α is 0·66 in monomer solution, then for such a solution $K' = 3{\cdot}68 \times 10^{-3}$.

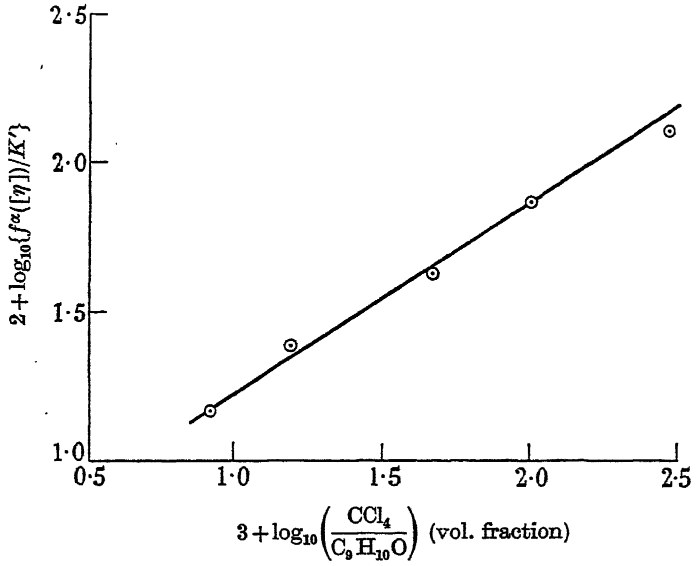


FIGURE 1. Polymerization in CCl_4.

TABLE 1

% Cl	m.w. (from Cl)	$[\eta]$	$K' \times 10^3$
0·225	62960	3·713	2·61
0·24	59030	5·781	4·10
0·82	17170	2·290	3·67
1·25	11220	1·323	2·81
2·81	4910	0·759	2·79
4·01	3400	0·635	2·97
5·78	2320	0·545	3·28

The mean value of K' is $3{\cdot}17 \times 10^{-3}$.

This value of α is confirmed by the variation of $d\eta/dt$ with light intensity. At high light intensities (3) reduces to $d\eta/dt = \text{constant}\, I^{\frac{1}{2}(1-\alpha)}$. The gradient of the plot of $\log d\eta/dt$ against $\log I$ for the highest intensities gives the value 0·18 for $\frac{1}{2}(1-\alpha)$. Hence $\alpha = 0{\cdot}64$, in good agreement with the value obtained independently above.

For solutions of polystyrene in toluene or styrene, K' and α have the values $4{\cdot}57\times10^{-3}$ and 0·65 respectively. Equation (1) may be written in the form $[\eta] = C\bar{P}^{\alpha}$, where $C = K'M_0^{\alpha}$. Then for solutions of poly-p-methoxystyrene and polystyrene in their monomers, $C = 9{\cdot}33\times10^{-2}$ and $9{\cdot}35\times10^{-2}$ respectively. This suggests that the intrinsic viscosity of substituted polystyrenes depends mainly on the length of the chains in solutions in which the solvent molecules resemble the polymer segments. It is interesting that the intrinsic viscosity of polystyrene is the same in toluene and monomer, while that of poly-p-methoxystyrene differs. Obviously toluene resembles styrene more closely than it resembles p-methoxystyrene.

It was shown in part I that the calculated velocity constants are not very sensitive to the value of α, provided that K' is chosen so as to give the correct molecular weight for the thermal polymer. Since only rational values of α can be used in calculating the after-effect (see part I), the values $\alpha = \frac{2}{3}$ and $K' = 3{\cdot}44\times10^{-3}$ have been adopted.

(c) *Measurements of* $\bar{P}$

The values of $\bar{P}$ for thermal polymer prepared at 30 and 60° C were determined by the viscosity method. The thermal polymer contains both initial and transfer polymer, and its number average degree of polymerization differs from its viscosity average. A correction must therefore be applied to the value of $\bar{P}'$ found from equation (1) by using the relation

$$\bar{P}^{\alpha} = \frac{2^{\alpha}(\beta+1)^{1+\alpha}}{(2\beta+1)^{\alpha}(\beta+1+\frac{1}{2}\alpha)}\bar{P}'^{\alpha}.$$

This correction was negligibly small in the case of polystyrene, because β was always large. Here, however, $\beta = 2{\cdot}3$ at 0° C, and it must decrease with increasing temperature. For reasons to be discussed later, the overall activation energy of β may be assumed to be -6 kcal. Hence $\beta_{30^\circ} = 0{\cdot}78$, and $\beta_{60^\circ} = 0{\cdot}32$. The corrected values for $\bar{P}$ and $\bar{M}$ obtained, using these values of β, are given in table 2. Since the production of thermal polymer at 0° C is inconveniently slow, a value for $\bar{P}$ at 0° C has been extrapolated from the values at 60 and 30° C. Taking $E_{\beta} = -6$ kcal., the value of $\bar{P}$ at 0° is calculated to be $4{\cdot}16\times10^4$, corresponding to $\bar{M} = 5{\cdot}58\times10^6$.

TABLE 2

° C	$[\eta]$	$\bar{P}$	$\bar{M}$
30	37·4	$1{\cdot}203\times10^4$	$1{\cdot}61\times10^6$
60	16·1	$3{\cdot}568\times10^3$	$4{\cdot}78\times10^5$

(d) *Measurement of the rate of polymerization*

The viscometers, of fused silica, were of the modified type described in part III (Bamford & Dewar 1949). All measurements were made in a laboratory lit with sodium lamps.

In the photopolymerization it is important to use light which is only weakly absorbed by the monomer (see part I). The absorption spectrum of p-methoxystyrene was therefore determined (in hexane solution) and is shown in figure 2. The extinction coefficient at 3650Å is approximately 0·05, a suitable upper limit. A medium-pressure mercury arc was used as source in the photopolymerization, and wave-lengths less than 3650Å were removed by soda-glass filters. Measurement of their extinction showed that under these conditions absorption by the monomer of shorter wave-lengths was negligible. All extinctions were measured with a Hilger Spekker photometer and medium quartz spectrograph.

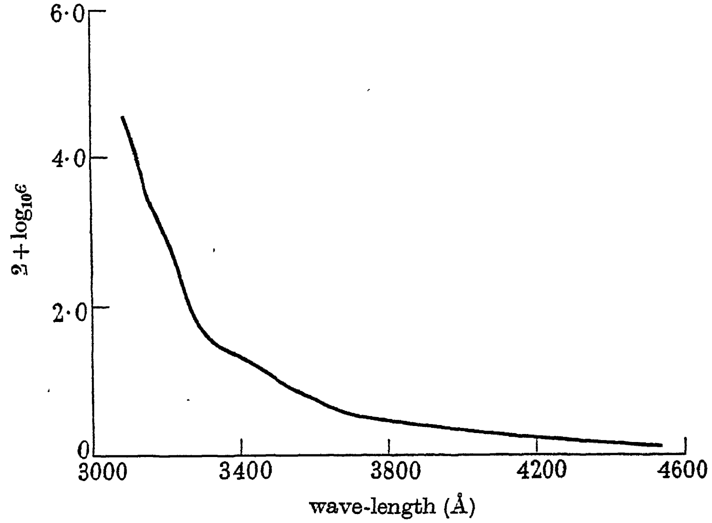


FIGURE 2. Absorption spectrum for *p*-methoxystyrene.

The reaction was followed only for very small conversions (see part I). Photochemical rates were measured at 0 and 25° C. At 25° C, varying the light intensity by a factor of 5000 produced a rate only four times greater than the thermal one at the same temperature. Such a variation is insufficient to enable the intensity-rate curve to be fitted to equation (3), and this, together with the fact that the after-effect at this temperature was very small, made it impossible to measure the velocity constants at 25° C. However, the results at 0° were generally satisfactory. They were fitted to equation (15), and values of A, β, and $(k_2^2/k_4)^{\frac{1}{2}} k_1^{\frac{1}{2}}$ thus obtained.

The calculated and experimental rates at 0° C are shown in table 3.

TABLE 3

Temperature 0° C, $A = 3{\cdot}5$, $\beta = 2{\cdot}3$

I (arbitrary units)	$10^6 d\eta/dt$ (sec.$^{-1}$) calc.	$10^6 d\eta/dt$ (sec.$^{-1}$) obs.	$\Delta\eta_I^{\infty} - \Delta'\eta_I^{\infty}$ (obs.)
0	0·420	0·420	
1·1	0·794	0·769	
9·8	1·53	1·44	
50·0	2·31	2·72	
775·0	4·05	4·33	0·0026
7740·0	6·09	5·91	0·0026

Measurement of the after-effect

If the monomer is irradiated for t_1 sec. and the viscosity measured after a further t_2 sec., the total increase $\delta\eta$ in the viscosity from the beginning of the irradiation may be written

$$\delta\eta = at_1 + bt_2 + \Delta\eta_I^{t_2} - \Delta'\eta_I^{t_1},$$

where a and b are the photochemical and thermal rates respectively, $\Delta\eta_I^{t_2}$ is the photochemical after-effect, and $\Delta'\eta_I^{t_1}$ is the pre-effect. The determination of $\delta\eta - bt_2$ for two or more values of t_1 enables $\Delta\eta_I^{t_2} - \Delta'\eta_I^{t_1}$ to be found. Actually, measurements were made when t_2 was so large that the rate had virtually fallen to its limiting value—this took about 2 hr. at 0° C. The quantity then measured is $\Delta\eta_I^{\infty} - \Delta\eta_I'^{\infty}$.

The after-effect at 25° C was too small to measure, and at 0° C it was much smaller than in the case of styrene. The values of the quantity $\Delta\eta_I^{\infty} - \Delta'\eta_I^{\infty}$ were not very reproducible at low intensities, and the constants were calculated from the value 0·0026 obtained at the highest intensities (table 3); the experimental error is estimated to be ± 20 %.

Results and discussion

(i) In table 4 the velocity constants, $\bar{P}$ and β, are given for *p*-methoxystyrene at 0° C. The corresponding values for styrene are also listed for comparison. Activation energies for *p*-methoxystyrene could not be estimated. As mentioned above, it was impossible to work at temperatures much above 0° C, while the freezing-point of the monomer is only slightly below 0° C. Even the measurements at 0° C are too inaccurate for this purpose.

Table 4

Temperature 0° C

	p-methoxystyrene	styrene
k_1	$5{\cdot}94 \times 10^{-16}$	$4{\cdot}51 \times 10^{-18}$
k_2	2·92	6·91
k_3	$5{\cdot}78 \times 10^{-5}$	$7{\cdot}47 \times 10^{-5}$
k_4	$1{\cdot}06 \times 10^{6}$	$1{\cdot}83 \times 10^{6}$
$\bar{P}$	$4{\cdot}16 \times 10^{4}$	$9{\cdot}08 \times 10^{4}$
β	2·3	26

(ii) In section (*c*), the activation energy of β was taken to be − 6 kcal. The general features of the polymerization suggest that the velocity constants do not differ widely from those of styrene with the exception of k_1, which is much larger for *p*-methoxystyrene. Thus $E_\beta = E_3 - \frac{1}{2}(E_1 + E_4)$ will probably be greater algebraically for *p*-methoxystyrene than for styrene, since E_3 and E_4 will be about the same for the two monomers while E_1 will be less for the former. It was assumed above that $E_\beta = -6$ kcal., which seems a reasonable estimate. For calculating $\bar{P}$ at 0° C, the exact value of E_β is not important, since the extrapolation is relatively insensitive to E_β.

(iii) It is possible to estimate E_1 approximately from the relation

$$(2\beta + 1)\,k_1[M]^2 = -\frac{1}{\bar{P}}\frac{d[M]}{dt},$$

together with the observed values for the rate of disappearance of monomer (table 5). These values lead to an activation energy of 30 kcal. for k_1. This figure is only reliable to about ± 2 kcal. in view of the uncertainty in E_β; it does, however, indicate definitely that the activation energy of initiation is much lower with *p*-methoxystyrene than with styrene, as was assumed above. (It leads, moreover, to a frequency factor 4×10^8, and it would seem that, for *p*-methoxystyrene, the substituent decreases both the energy and the entropy of activation of the initiation process.)

TABLE 5

temperature (° C)	$-\frac{d[M]}{dt}$ (mol.l.$^{-1}$sec.$^{-1}$)	$-\frac{1}{(2\beta+1)\bar{P}}\frac{d[M]}{dt}$ (mol.l.$^{-1}$sec.$^{-1}$)
0	$3{\cdot}94 \times 10^{-9}$	$1{\cdot}69 \times 10^{-14}$
30	$1{\cdot}22 \times 10^{-7}$	$3{\cdot}97 \times 10^{-12}$
60	$2{\cdot}01 \times 10^{-6}$	$3{\cdot}44 \times 10^{-10}$

(iv) Table 6 shows the calculated lifetimes of a single radical, a growing polymer molecule, and a kinetic chain, in the thermal polymerization. Evidently the radical derived from *p*-methoxystyrene is appreciably more stable than that from styrene. The *p*-methoxy substituent would be expected to increase the resonance energy of the polystyryl radical, and so to increase k_1, and to lower the other three velocity constants.

TABLE 6. MEAN LIFETIMES AT 0° C

radical	$4{\cdot}54 \times 10^{-2}$ sec.
polymer molecule	32·0 min.
kinetic chain	1·46 hr.

Walling, Briggs, Wolfstirn & Mayo (1948) give monomer reactivity ratios for the co-polymerization of styrene and *p*-methoxystyrene. From these, and the values for k_2 in table 4, the absolute velocity constants for the crossed propagations may be calculated (table 7).

TABLE 7

	monomer		
radical	styrene	*p*-methoxystyrene	ratio
styrene	6·91	5·95	1·16
p-methoxystyrene	3·56	2·92	1·22
ratio	1·95	2·04	

From this table, the following conclusions can be drawn:

(1) Towards a given radical, *p*-methoxystyrene monomer is less reactive than styrene monomer.

(2) Towards a given monomer, *p*-methoxystyrene radicals are less reactive than styryl radicals.

(3) The difference in reactivities is much greater for the radicals than for the monomers.

Since the resonance energy of the radicals will be greater than that of the monomers, the effect of the substituent in changing it would be expected to be greater, in accordance with conclusion (3) above. Shorygin & Shorygina (1939) have studied the effect of substituents on what they describe as the 'polymerization capacity' of styrene. Their quantitative deductions are vitiated by their use of the simple

Staudinger relation to determine degrees of polymerization, but they do predict a decreased polymerization capacity in *p*-methoxystyrene as compared to styrene. This agrees with the present results.

(v) The results obtained here confirm the reaction scheme put forward in part I of this series. It is pointed out in part III (Bamford & Dewar 1949) that, if termination occurred by combination, then the rate of viscosity increase in the photopolymerization would be proportional to $\sqrt{I}$ at high light intensities. In fact, $d\eta/dt$ is proportional to $I^{0\cdot18}$ under these conditions, and therefore termination cannot occur by combination.

The author is grateful to Dr C. H. Bamford and Dr M. J. S. Dewar for their interest in this work.

References

Bamford, C. H. & Dewar, M. J. S. Part I 1948*a* *Proc. Roy. Soc.* A, **192**, 309.
Bamford, C. H. & Dewar, M. J. S. Part II 1948*b* *Proc. Roy. Soc.* A, **192**, 329.
Bamford, C. H. & Dewar, M. J. S. Part III 1949 *Proc. Roy. Soc.* **197**, 356.
Frank, R. L. & Smith, P. V. 1946 *J. Amer. Chem. Soc.* **68**, 1365.
Houwink, R. 1940 *J. prakt. Chem.* **157**, 15.
Johnson, J. R. 1942 *Organic reactions*, I. New York: Wiley and Sons.
Mayo, F. R. 1943 *J. Amer. Chem. Soc.* **65**, 2324.
Schulz, G. V. & Sing, G. 1943 *J. prakt. Chem.* **161**, 161.
Shorygin, P. P. & Shorygina, N. V. 1939 *J. Gen. Chem., Moscow*, **9**, 845.
Walling, C., Briggs, E. R., Wolfstirn, K. B. & Mayo, F. R. 1948 *J. Amer. Chem. Soc.* **70**, 1537.
Walling, C. & Wolfstirn, K. B. 1947 *J. Amer. Chem. Soc.* **69**, 852.

Evaporation and surface structure of liquids

By G. Wyllie, *H. H. Wills Physical Laboratory, University of Bristol*

(*Communicated by N. F. Mott, F.R.S.—Received* 29 *October* 1948)

The condensation coefficient (c.c.) of a liquid is defined as the ratio of the observed rate of evaporation under given conditions to the 'expected' rate calculated from the known saturation vapour pressure, for the same conditions. Values of the c.c. less than 1 have been reported in the past, but the accuracy of the experiments has been criticized on various grounds. Experimental work is described which establishes that the c.c. of glycerol at 18° C is $0{\cdot}052 \pm 0{\cdot}005$. Previous experimental determinations are reviewed in the light of this result, and it is found that for various liquids the c.c. is nearly equal to the free-angle ratio defined by Kincaid and Eyring on the basis of a simple model of liquid structure. This equality is not found for chloroform.

The theory of evaporation is discussed on a model of the liquid surface consisting of a compact ordered layer with a mobile adsorbed layer of molecules above it. On the assumption that evaporating and condensing molecules pass through the adsorbed condition, it is shown that the c.c. is determined by the rates of exchange between substrate and adsorbed layer, and adsorbed layer and vapour, respectively. In particular, a c.c. near unity results if the rate of exchange between substrate and mobile layer is very much greater than that between mobile layer and vapour. The above assumption does not give a satisfactory explanation of the equality of c.c. and free-angle ratio for alcohol molecules.

It is shown that for polar molecules which form an oriented surface layer in the liquid state it is to be expected that evaporation will take place mainly by direct exchange of molecules between the ordered layer and the vapour, although an adsorbed layer will be present. In this case an argument given by Herzfeld shows the condition for equality of c.c. and free-angle ratio to be that there should be no exchange of energy between rotational and translational co-ordinates in the process of evaporation. It is further concluded from a rather qualitative discussion that a major contribution to c.c. and free-angle ratio for the simple alcohols arises from restriction of rotation of the hydroxyl group about the carbon-oxygen bond.

Introduction

The maximum possible number of molecules leaving a unit area of the surface of a substance per unit time when it is in equilibrium with its vapour is equal to the number striking a unit area per unit time from the vapour. The latter is given by the kinetic theory, and is

$$(2\pi mkT)^{-\frac{1}{2}}p_0 \text{ molecules cm.}^{-2}\text{sec.}^{-1},$$

where p_0 is the saturation vapour pressure at temperature T, and m the mass of a molecule.

If now the space over the surface be kept at a pressure of vapour p less than p_0, we may calculate an expected rate for the evaporation assuming that the change in density of the vapour has no effect on the rate at which molecules leave the condensed phase. Since the latent heat of expansion of the vapour is so much less than the latent heat of vaporization of the condensate, this assumption seems highly plausible. The expected rate is then

$$(2\pi mkT)^{-\frac{1}{2}}(p_0 - p) \text{ molecules cm.}^{-2}\text{ sec.}^{-1}$$

It is found, however, that for a number of materials the rate of evaporation determined experimentally is very much less than the expected value.

These experimental results may be accounted for in various ways, as follows:

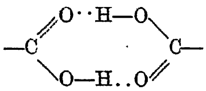
(*a*) There may be a considerable probability of a molecule arriving from the vapour being at once reflected without entering the condensed phase.

(*b*) Molecules may condense on the surface in a configuration other than that of the underlying material and subsequently either re-evaporate or become built in to the condensed structure.

(*c*) The molecular species involved may be wrongly identified, as in the case of phosphorus discussed below.

In any of these cases we may define a quantity, the condensation coefficient, as the ratio of the actual to the expected rate of evaporation. This may or may not be independent of the pressure over the surface. It measures the probability that a molecule arriving from the vapour will condense in the strict sense, i.e. in such a way as to form part of the structure of the condensed phase.

Melville & Gray (1936*a*, *b*) have shown that a discrepancy of 10^{-7} in the rate of evaporation of red phosphorus is due to the fact that in an enclosed space red phosphorus reaches equilibrium with a vapour of P_4 molecules, whereas the units taking part in evaporation and condensation are P_2 molecules. Again, Haward (1939) has observed that the condensation of benzoic acid from a rough molecular stream on to a deeply cooled surface involves the formation of dimers in a mobile adsorbed film on the crystal surface. The importance of this effect depends on the saturation of intermolecular forces in small groups—pairs in each of the cases mentioned—corresponding either to primary valence bond or hydrogen-bond formation. The carboxylic acids, which dimerize with the formation of a ring



and are known to be paired in this way in the solid state, accordingly show this behaviour in a particularly simple form.

Langmuir (1916) has pointed out that molecules of the vapour approaching the condensate surface are attracted, and so will arrive at the surface with average energy in excess of that corresponding to the temperature. Their interaction with the surface molecules will in general result in the transfer of this excess energy to the molecules of the condensate, so that most incident molecules will remain more or less loosely bound to the surface until they are again thermally excited. Detailed calculation has not so far proved practicable save for rather simplified cases (see Alty, 1937) and the accommodation coefficient, although it has been measured for various gases and vapours on metal surfaces, has not been found for a vapour on its own condensate. There is thus no conclusive evidence of the importance of possibility (*a*).

For solid crystals, alternative (*b*) is now clearly established by the work especially of Volmer (1939) and his collaborators, and by the investigations of Cockcroft (1928), Haward (1939) and Neumann & Costeanu (1939). However, the experimental work for solids is complicated by several factors:

(i) In polycrystalline material the surface area is quite uncertain.

(ii) Even for single crystals the condensation coefficient depends on the crystal face considered.

(iii) Unless the crystal has its equilibrium form, etch pits, i.e. hollows in one face bounded by other crystal planes, are likely to develop.

For liquids, on the other hand, the surface can safely be taken to be the apparent superficies (unless there is a likelihood of film-spreading such as spoiled the attempt of Neumann & Costeanu to find the condensation coefficient of liquid potassium perrhenate). Further, the surface rapidly adjusts itself to changed conditions, attaining an equilibrium structure which then remains constant until conditions are again changed. These advantages make determination of the condensation coefficient for liquids of interest in spite of the difficulty of relating the results to an imperfectly understood bulk structure.

Prüger (1940) gives a review of previous work on this subject and shows that discordant results are generally to be ascribed to the difficulty of obtaining measurements of the temperature of the evaporating surface. His own careful measurements on water and carbon tetrachloride near their respective boiling-points, using a small pressure defect over the surface and determining the surface temperature directly by means of a fine thermojunction, give results agreeing within a factor of 2 with those obtained by Alty & Nicoll (1931), Alty (1933), and Alty & Mackay (1935). Alty in his latest work on water measured surface tension to obtain temperature, using lower vapour pressures but a relatively much larger pressure defect than Prüger. Baranaev (1939) made a series of determinations at about room temperature using very accurate pressure measurement and a small pressure defect over the surface. However, all these experiments at rather high pressures—several mm. of mercury and upwards—are open to two objections. On the one hand, it is conceivable that the rate of evaporation does depend very much on the pressure of vapour actually present, though if this were the case the substantial agreement of Alty's and Prüger's results for water is surprising; on the other, the large temperature gradient at the liquid surface associated with the rather large rate of evaporation makes the measurement of surface temperature uncertain. This last criticism is less substantial for Baranaev's work than for the others', but unfortunately his thermocouple seems to have been less refined than that used by Prüger, so that the small drop in temperature that he found at the surface cannot be regarded as established by the measurement.

The only previous experimental work which avoids both of these criticisms is that of Knudsen (1915) on mercury. His finding that the condensation coefficient of mercury vapour on a clean surface of liquid mercury is unity, however, leaves open the question of the value for materials consisting of less simple molecules. It thus became evidently desirable to determine whether a liquid of low vapour pressure could be found to have a condensation coefficient of less than unity under similar conditions.

EXPERIMENTAL

It was observed that all those substances, which in previous work had shown condensation coefficients markedly less than 1, had large dipole moments. Accordingly, in the present work various polar substances of fairly large molecular weight were examined. Glycerol was finally selected for precise measurement. The material used was 'AnalaR' reagent thoroughly dried but not otherwise purified.

The experimental method adopted was to determine the vapour pressure by the molecular effusion method, then to remove the cap carrying the effusion orifice and find the rate of evaporation from the same vessel into vacuum.

The arrangement of the apparatus is as follows. An aluminium standard of 10 cm. overall length was ground into a pyrex tube of 2 cm. internal diameter, and had its upper end recessed to carry the experimental vessel. The tube was connected through a liquid-air trap to a vacuum system. A thermocouple was clamped to the head of the standard and the wires led out through the base, where they were sealed with Apiezon hard wax.

The first design of the apparatus incorporated a glass surface cooled by liquid air, 1 cm. from the cap of the experimental vessel. It was found, however, that loss of heat by radiation caused a temperature gradient in the walls of the vessel sufficient to make it act as a reflux condenser, so the arrangement shown in figure 1 was substituted. The vessel was initially made from stainless steel, but the liquids used in the experiments spread over the surface of the metal, making the area of the evaporating surface quite indeterminate, so a thin glass tube ground into a brass cap as in figure 1 was used instead, the recess in the head of the standard being lined with the same vacuum grease as was used for the joints of the apparatus, to ensure satisfactory thermal contact between the aluminium and the glass. The glass had the advantage that it could be very thoroughly cleaned by washing with various agents and then heating to the annealing temperature in a blow-pipe flame—an operation of some difficulty with such thin material.

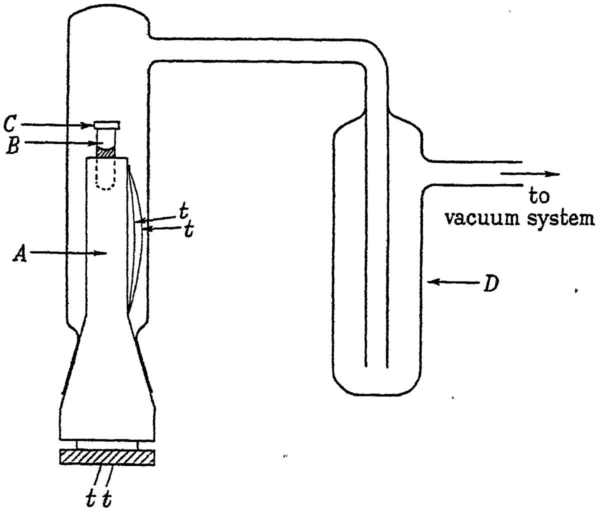


FIGURE 1. Arrangement of apparatus. *tt*, thermocouple leads; *A*, aluminium standard; *B*. glass tube containing liquid; *C*, brass cap carrying foil with effusion hole; *D*, liquid-air trap.

The effusion hole was drilled in thin platinum foil. The least ragged hole was selected and the platinum soldered to the brass cap.

The experimental procedure was as follows. A clean glass tube containing about 5 g. of glycerol was put into the apparatus in place of the standard, with a glass stopper supporting the tube in the ground joint, and kept under vacuum for a week to remove the last traces of water. A few drops were transferred to the experimental vessel, care being taken to avoid as far as possible the transfer of material from the surface layer. The vessel was weighed, mounted in the standard, and inserted in the apparatus.

Since an oil-diffusion pump was used to obtain high vacuum, the apparatus was first pumped out with the rotary pump, filled with hydrogen at about half atmospheric pressure, and cut off from the vacuum system while the oil-diffusion pump was brought into action. Then the apparatus was re-connected to the vacuum system. Liquid air was applied to the trap when the pressure fell below 10^{-4} mm. Hg, and the experiment was timed from that moment.

A series of experiments with indene showed that 5 min. should be added to the time of an experiment to allow for this preliminary manipulation. At the end of an experiment the apparatus was cut off from the pumps, air admitted and the experimental vessel at once removed and weighed. Vapour pressure and free evaporation experiments were conducted alternately, the duration of an experiment varying from 5 to 35 hr.

During the experiments on glycerol the temperature of the apparatus was kept constant by a thermostatically controlled air-bath. The temperature indicated by the thermocouple remained within the limits 17·9 and 18·1° C while the bath temperature fluctuated between 17·5 and 18·5° C, the aluminium standard proving quite a good thermal shock-absorber.

The preliminary experiments on indene indicated that this material had too high a vapour pressure (0·6 mm. Hg at 15° C) and evaporated much too rapidly for this method to be applicable with any accuracy. The observations showed, however, that the condensation coefficient certainly did not differ from unity by a factor of so much as 3.

With glycerol the rate of evaporation was very slow both with the cap on and off the vessel. There was no difficulty in determining surface area, since the glycerol failed to wet the freshly annealed glass surface, giving a contact angle of about 30°. Rough preliminary experiments indicated a surprisingly low value (10^{-3} to 10^{-2}) for the condensation coefficient.

Such low results, if spurious, may be accounted for in two ways. Either the liquid surface may be contaminated or the vapour pressure may be overestimated. All possible precautions were taken to avoid the first of these, but it still seemed possible that vapour from the grease in the apparatus might be affecting the surface. This would be expected to show itself in a steady decrease in the rate of free evaporation of a sample as it was kept in the vacuum. In fact, if a sample was taken which had not been dried in high vacuum in the apparatus, it did show a steady decrease by a factor of the order of 5. However, the time for the rate to fall to a given fraction ($\frac{1}{2}$) of its initial value was proportional to the mass of the sample—a result consistent with the removal of the last traces of dissolved water but not with the deposit of vapour on a surface which was independent of the mass.

Further, contamination of the surface by heavy molecules from the grease would be expected to have more affect on the rate of evaporation than on the vapour pressure and so give rise to a reduction of the condensation coefficient. The opposite effect was observed.

Overestimate of the vapour pressure might arise from leakage at the edge of the cap. This was easily eliminated by sealing the edge in a few experiments. No effect was observed.

Another source of error might be the escape of molecules by surface migration round the edge of the hole. However, when a smaller hole was substituted, the rate of evaporation was, within experimental error, proportional to the area of the hole, a result inconsistent with substantial escape by surface migration which would be proportional to the perimeter.

Finally, the vapour pressure might be overestimated if the liquid were not adequately dried. The final results were obtained with material which had been kept *in vacuo* till its rate of free evaporation was unaltered by 3 days' further drying. The rather higher figure obtained for the condensation coefficient shows that the earlier results were in fact obtained with imperfectly dry material.

Results for the rates of evaporation were finally reproducible within the error of weighing. A certain amount of accuracy was sacrificed to expedition by the use of a chain device. This is the major contribution to the estimated error in the vapour pressure.

RESULTS

The measurements were interpreted using the formula: rate of efflux of vapour per unit area from an orifice smaller than the mean free path of the molecules in the vapour =

$$3{\cdot}45(M/T)^{\frac{1}{2}}\,P\ \text{g. min.}^{-1}\,\text{cm.}^{-2},$$

where M is the molecular weight in terms of the atom of hydrogen as unit (92 for glycerol) and P is in mm. Hg.

$$\text{Thus vapour pressure} = (1/3{\cdot}45)\,(T/M)^{\frac{1}{2}} \times \frac{\text{rate of evaporation from hole}}{\text{area of hole}}$$

$$= 0{\cdot}14 \pm 0{\cdot}01 \times 10^{-3}\,\text{mm. Hg at } 18^\circ\,\text{C.}$$

Again the condensation coefficient is given by

$$f = \frac{\text{rate of evaporation from free surface}}{\text{rate of evaporation from hole}} \times \frac{\text{area of hole}}{\text{area of free surface}}$$

$$= \frac{(1{\cdot}40 \pm 0{\cdot}03) \times 10^{-5}\,\text{mg. min.}^{-1}\,\text{cm.}^{-2}}{(2{\cdot}7 \pm 0{\cdot}2) \times 10^{-4}\,\text{mg. min.}^{-1}\,\text{cm.}^{-2}}.$$

$$= 5{\cdot}2 \pm 0{\cdot}5 \times 10^{-2}\ \text{at } 18^\circ\,\text{C.}$$

This result is calculated using the full superficial area of the free surface. Since the surface is concave there is a correction which in the limiting case when $f = 1$ reduces the effective surface to the cross-section of the tube. This correction, however, diminishes very rapidly with decreasing f, and is in this case quite negligible.

DISCUSSION

In the light of the above results it seems profitable to review the previously determined values of the condensation coefficients of liquids. These are summarized in table 1 along with other relevant data.

TABLE 1

liquid	temperature (° A)	measured c.c. (f)	free-angle ratio (δ)	Δs* (e.u.)
CCl_4†	273	1	1	20·7
C_6H_6†	279	0·90	$>0·85$	21·2
$CHCl_3$†	275	0·16	0·54	21·7
C_2H_5OH†	273	0·020	0·018	27·0
CH_3OH†	273	0·045	0·048	26·5
H_2O‡	283 to 303	0·036 to 0·040	0·04	27·3

* Δs is the entropy of vaporization at that temperature for which the density of saturated vapour is 0·0202 mole/l.

† Baranaev (1939).

‡ Alty (1933).

Eyring and his collaborators (Hirschfelder, Stevenson & Eyring 1937; Eyring & Kincaid 1938) have developed an account of liquid structure in which the partition function for the translatory motions of the molecules is expressed in terms of a free volume which can be estimated from the velocity of sound in the liquid. From the estimated free volume and the observed vapour pressure, Eyring & Kincaid evaluate the 'free-angle ratio', i.e. the ratio of the partition function for rotation of the molecules in the liquid to that for the molecules in the vapour, on the assumption that the partition function of internal vibration of a molecule is unaltered by the transition from the liquid to the vapour phase.

There is a surprisingly close agreement between the values of condensation coefficient and free-angle ratio for benzene, carbon tetrachloride and liquids which are associated by hydrogen bonds and have an anomalously high entropy of vaporization at constant normal volume of vapour. This agreement fails for chloroform which, in spite of its high dipole moment, does not form hydrogen bonds and has a relatively low anomaly in entropy of vaporization.*

It will be seen that we can interpret these observations in terms of the surface structure of the liquids.

In the disordered interior of a liquid, a molecule (or group of molecules) can arrange itself in quasi-crystalline grouping with its neighbours on one side or the other, but generally not both. Any regular order temporarily established is destroyed by thermal fluctuation in a time comparable with the period of molecular vibration. At the free surface, however, a molecule has neighbours on one side only and a much greater degree of order may be preserved. This is illustrated by the considerable electrical moment of the surface double layer in water, for example.

---

* Further, there is an interesting degree of agreement between the experimental results of Baranaev and the values calculated by him on the assumption of an energy of activation for evaporation, which he obtained by comparison of the surface free energy with the internal energy of vaporization. As Fuchs (1933) pointed out in connexion with a suggestion of Miyamoto, an activation energy such as this is regained by a molecule on escape, so that molecules evaporating from the surface of a liquid would have a mean velocity much higher than that corresponding to the temperature of the liquid. This effect has not been observed, though for low values of the condensation coefficient it should be considerable, so Baranaev's interpretation seems unlikely to be correct.

Thus we may consider as a first model for the structure of a liquid surface a regular layer not dissimilar, at least locally, to the solid surface; and loosely bound to it a layer of comparatively freely rotating mobile molecules such as has been shown to exist on many solid crystals. We shall analyze first the simplified case in which all molecules evaporate or condense by passing through the intermediate mobile layer.

Polanyi & Wigner (1928) have given a formula for the probability that a molecule will escape radially in a solid angle ω from a site of potential energy ϵ_w about which it oscillates with frequency ν over a potential hump ϵ_0 in time dt; this is

$$\frac{\omega\nu}{\pi}\frac{\epsilon_0-\epsilon_w}{kT}\mathrm{e}^{-(\epsilon_0-\epsilon_w)/kT}dt.$$

We may apply this formula to determine the rate of escape, considering only the translatory motion of the molecules. Since the long-range interaction between molecules due to van der Waals' forces is attractive in character, we should expect that there should be no extra activation energy required for this transition and that ϵ_0 should be the minimum potential energy of molecules in the mobile adsorbed layer. This condition can be relieved without materially altering the subsequent argument.

If the molecules can escape from sites of different types, and there are on the average N_r sites of type r on a unit area of surface, the number of molecules transferred to the mobile layer per unit of area and time is

$$\sum_r \frac{N_r\omega_r\nu_r}{\pi}\frac{\epsilon_0-\epsilon_r}{kT}\mathrm{e}^{-(\epsilon_0-\epsilon_r)/kT}.$$

Under equilibrium conditions this must be balanced by the number condensing out of the mobile layer. The number condensing to sites of type r per unit area per unit time must be of the form

$$\phi_r N_r n,$$

where n is the concentration of molecules in the mobile surface layer. If n_0 is the value of this concentration for a surface in equilibrium with the saturated vapour, we obtain

$$\phi_r n_0 = \frac{\omega_r\nu_r}{\pi}\frac{\epsilon_0-\epsilon_r}{kT}\mathrm{e}^{-(\epsilon_0-\epsilon_r)/kT}$$

for the value of ϕ_r at concentration n_0. We suppose that to a reasonable approximation ϕ_r is independent of n. If, further, we suppose N_r to be independent of n, the resultant rate of transfer of molecules into a mobile surface layer of concentration n is

$$\sum_r \phi_r N_r(n_0-n) = \sum_r \frac{\omega_r\nu_r N_r}{\pi}\frac{\epsilon_0-\epsilon_r}{kT}\mathrm{e}^{-(\epsilon_0-\epsilon_r)/kT}\frac{n_0-n}{n_0}.$$

Now let us consider the exchange of molecules between the vapour and the mobile layer. If the pressure of vapour over the surface is p, the number of vapour molecules falling into the mobile layer per unit area per unit time is $(2\pi mkT)^{-\frac{1}{2}}p$. If we take the zero of potential energy to be that of a vapour molecule infinitely far from the surface, the energy required for a molecule in the mobile layer to escape from its equilibrium distance from the surface to infinity is $-\epsilon_0$, so, as Frenkel has shown, the number of such molecules escaping per unit area per unit time is $n\nu_0 e^{\epsilon_0/kT}$.

At the saturation vapour pressure, then, we have for the rate of exchange between surface layer and vapour

$$n_0\nu_0 e^{\epsilon_0/kT} = (2\pi mkT)^{-\frac{1}{2}}p_0,$$

and for the exchange between surface layer and substrate

$$\sum_r \phi_r N_r n_0 = \sum_r \frac{\omega_r \nu_r N_r}{\pi} \frac{\epsilon_0 - \epsilon_r}{kT} e^{-(\epsilon_0-\epsilon_r)/kT}.$$

We notice that there is no necessity for these rates of exchange to be equal.

Now consider the state of steady evaporation or condensation in which p is maintained at a steady value not equal to p_0. Then the net rate of transfer of molecules from the surface layer into the vapour is

$$n\nu_0 e^{\epsilon_0/kT} - (2\pi mkT)^{-\frac{1}{2}}p,$$

which must be the same as the net rate of transfer from the substrate into the surface layer

$$\sum_r \phi_r N_r(n_0 - n),$$

so that

$$n(\nu_0 e^{\epsilon_0/kT} + \sum_r \phi_r N_r) = \sum_r \phi_r N_r n_0 + (2\pi mkT)^{-\frac{1}{2}}p,$$

i.e.

$$\frac{n}{n_0} = \frac{a+bp}{a+bp_0},$$

where

$$a = \sum_r \phi_r N_r n_0 \quad \text{and} \quad bp_0 = (2\pi mkT)^{-\frac{1}{2}}p_0,$$

are the rates of exchange between surface layer and substrate and between surface layer and vapour at equilibrium, respectively. Thus the net rate of evaporation is

$$\frac{a+bp}{a+bp_0}bp_0 - bp,$$

$$= \frac{a}{a+bp_0}b(p_0-p),$$

and comparing this with the experimental definition of the condensation coefficient we obtain

$$f = \frac{a}{a+bp_0}.$$

In other words, if the rate of transfer of molecules from the substrate into the surface layer is so large that the concentration in that layer is sensibly independent of the pressure of vapour over it, the condensation coefficient will be unity; if, on the other hand, that rate is small, the condensation coefficient will be nearly its ratio to the rate of exchange between surface layer and vapour at equilibrium.

So far this argument has taken no account of the rotations of the molecules. Herzfeld (1935) has shown that the rate at which molecules escape from a potential well is very much dependent on whether there is exchange of energy between the rotational and translational degrees of freedom. In fact, for molecules in such a field that their vibrations are simple harmonic, if the energy required to remove a molecule to infinity from a given position of its centre of gravity is decreased below its value for the equilibrium position by the potential energy of rotational as well as translational displacement, the rate of escape is increased by the reciprocal of the free-angle ratio.

Now it is evident that if the rotation of molecules is restricted in the condensed phase, but not in the adsorbed layer or the vapour, the concentration of molecules in these latter regions is increased, by the reciprocal of the free-angle ratio, beyond what would be expected on the argument given above, which neglects rotation. If the potential energy of rotation does not contribute to the escape of a molecule from the substrate, however, the rate of exchange between substrate and adsorbed layer is unaltered in spite of the change of concentration, so that for the condensation coefficient we must replace $f = a/(a+bp_0)$ by $f = a/(a+bp_0/\delta)$. If, in accord with the experimental data, this is put equal to δ, then if δ is small we must have $a \doteqdot bp_0$, which we have no reason to expect. This, moreover, would imply that in the absence of the hindrance to rotation in the liquid the condensation coefficient would be 0·5 instead of 1, so it appears that on our assumptions restricted rotation by itself is not adequate to account for the observed condensation coefficient.

This discussion, however, has all been conducted on the supposition that condensation and evaporation proceed through the intermediate mobile adsorbed layer. If we can show reason why this should not occur, Herzfeld's argument can be at once applied to the direct exchange of molecules between vapour and condensate.

Such a reason can be found for those polar molecules which form a more or less uniformly oriented layer at the surface. The resulting sheet of dipoles has, at distances large compared with the intermolecular distance, the same effect as a uniform electrical double layer. The removal of one molecule from the sheet—that is, the formation of a site into which a molecule could condense—gives rise to a field which is that of an oppositely oriented dipole at the vacant site, and consequently repels a molecule which is correctly oriented for condensation and is within the narrow range of solid angle required. A molecule of correct orientation some distance off to the side will experience an attraction as indicated in figure 2.

This repulsive effect is superposed on the normal attractive field, the principal part of which arises from the high polarizability of the liquid and is simply the electrostatic image force (very little less than for a conductor). The forming of a site for condensation is then equivalent to depolarizing the liquid locally, so that the potential energy of a molecule properly oriented for condensation in the neighbourhood of a vacant surface site is as shown in the contour diagram of figure 2.

The net result is that there is no activation energy for the exchange of molecules between liquid and vapour, but a substantial one for the exchange between liquid and adsorbed layer. In equilibrium this is counterbalanced by the fact that the concentration of molecules in the adsorbed layer is higher than in the vapour by a Boltzmann factor of approximately the same magnitude as that arising from the activation energy for complete condensation, so that the ratio of direct evaporation to evaporation through adsorbed layer at equilibrium is determined by a solid angle factor only. Consequently, the total rate of evaporation is determined within a factor of the order of 2 by the direct evaporation, when the pressure of vapour is less than the saturated vapour pressure.

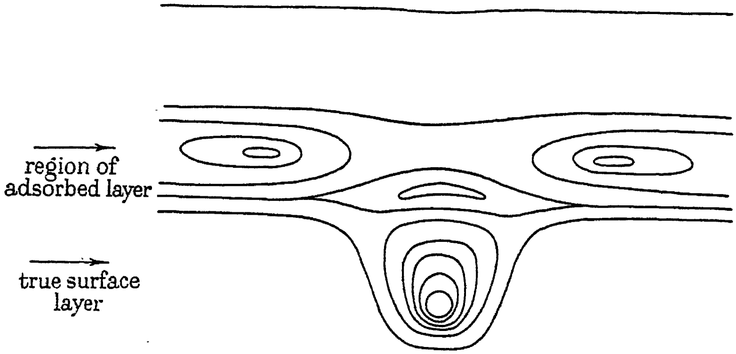


FIGURE 2. Potential energy contours for a molecule properly oriented for condensation in the neighbourhood of a vacant surface site.

It is further probable that the rate of exchange between substrate and adsorbed layer is, in the range of measurement, dependent on the concentration in the adsorbed layer. In equilibrium this concentration must be rather high, as the electrostatic binding of an adsorbed molecule is a considerable fraction of the heat of evaporation. The dipole interaction of these molecules is long-range, so the layer cannot reasonably be described as an ideal two-dimensional gas. This is confirmed by Alty's results, for since the surface temperature and the corresponding vapour pressure are very sensitive to small variations in surface tension, the good agreement of his result for the condensation coefficient, finding surface temperature by surface tension, with that obtained by a rougher but more direct method indicates that the surface pressure of the mobile adsorbed layer, whose concentration is much reduced under the reduced vapour pressure, is lower than would correspond to an ideal case. Thus in evaporation the direct exchange between vapour and condensate, and in condensation the indirect exchange, will be relatively more important than at equilibrium. The importance of this effect could be found by measurement of the condensation coefficient both in condensation and evaporation, but the writer is not aware of any such measurements at constant surface temperature.

Then the observed equality of condensation coefficient and free-angle ratio for hydrogen-bonded liquids can be explained by the supposition that the kinetic energy required for a molecule to escape from its position in the surface is independent of tilt—when tilt refers to whatever co-ordinates are in fact involved in determining the free-angle ratio. This is obviously untrue for the rigid rotations of the molecule, considering only electrostatic forces, and it is unlikely that its neighbours should so jam a tilted molecule as just to verify the supposition: however, for these materials, Eyring & Kincaid's assumption that the internal motions of the molecules are the same in the vapour and the condensed phase is certainly untrue. The motion of the hydroxyl hydrogen atom of an alcohol molecule relative to its neighbour oxygen is completely altered by the hydrogen-bond formation, and this may give the major contribution to the apparent free-angle ratio.

It is necessary, first, to consider the fact that in the hydrogen bond the hydrogen atom concerned has two alternative asymmetric positions between the two oxygen atoms. The contribution to the entropy of the condensed phase corresponding to this can be regarded as arising from a modification of that internal vibration of the molecule in which the bonding hydrogen moves towards and away from the oxygen. This will give an increase of the free-angle ratio deduced from measurement, by a factor of 2 for completely random distribution of the protons.

On the other hand, if we consider the possible arrangements of bonding hydrogen atoms in any particular chain of bonds it is evident that only a few arrangements do not have high electrostatic energy. Thus, although the exchange of hydrogen between neighbouring oxygen atoms is of great importance to the dielectric properties, the corresponding factor in the partition function is small in effect compared with that arising from restricted rotation of the hydrogen around the oxygen atom.

It is possible to estimate the frequency of transverse vibration of the proton in a hydrogen bond by considering it sliding on the surface of one oxygen ion under the field of the other. An estimate of the minimum effective charge on hydrogen and oxygen may be obtained from the known dipole moment and size of the normal hydroxyl group. The frequency thus estimated is of the order of $1{\cdot}5 \times 10^{13}$ sec.$^{-1}$, so the vibration is highly degenerate at room temperature

$$(h\nu \sim 10 \times 10^{-14}, \quad kT \sim 4 \times 10^{-14} \text{ erg})$$

and the corresponding factor in the partition function is near unity. On the other hand, the partition function of the free* rotation of the OH group about the C—O bond is 6, so that the greater part of the 'free-angle ratio' for, e.g. methyl alcohol may arise from this source.

Each water molecule gives hydrogen to two hydrogen bonds in the crystalline state, so the whole of the free-angle ratio for water is directly due to this effect. For the alcohols, the residual part of the free-angle ratio is due to the interaction of the paraffin tails coupled with the comparative rigidity of the H—O—C angle. This probably accounts for the difference between methyl and ethyl alcohol.

Thus it is possible to account at least qualitatively for the observed equality of condensation coefficient and free-angle ratio in water and the alcohols, in terms of a simple model of the surface structure and the known properties of the hydrogen bond. For such liquids as benzene or carbon tetrachloride, it is more doubtful whether the model is physically valid, but the high condensation coefficient observed may be explained by saying that interchange of molecules between the two surface layers is much more rapid than any other process in the evaporation.

---

* There is evidence from the dielectric properties and molecular spectra that in the vapour this rotation is effectively free. See Onsager (1936) and Herzberg (1945, p. 497).

The author wishes to record his gratitude to Professor T. Alty, under whose direction the experimental investigation was carried out, to the Senate of Glasgow University for facilities for the work, and to Professor N. F. Mott, F.R.S., for discussions of the theoretical aspects.

REFERENCES

Alty, T. 1933 *Phil. Mag.* **15**, 82.
Alty, T. 1937 *Sci. Progr. Twent. Cent.* **123**, 436.
Alty, T. & Mackay, C. A. 1935 *Proc. Roy. Soc.* A, **149**, 104.
Alty, T. & Nicoll, F. H. 1931 *Canad. J. Res.* **4**, 547.
Baranaev, M. 1939 *J. Phys. Chem. U.S.S.R.* **13**, 1635.
Cockcroft, J. D. 1928 *Proc. Roy. Soc.* A, **119**, 293.
Eyring, H. & Kincaid, J. F. 1938 *J. Chem. Phys.* **6**, 620.
Fuchs, N. 1933 *Phys. Z. Sowjet.* **4**, 481.
Haward, R. N. 1939 *Trans. Faraday Soc.* **35**, 1401.
Herzberg, G. 1945 *Infrared and Raman spectra of polyatomic molecules.* New York: D. van Nostrand.
Herzfeld, K. 1935 *J. Chem. Phys.* **3**, 319.
Hirschfelder, J. O., Stevenson, D. P. & Eyring, H. 1937 *J. Chem. Phys.* **5**, 896.
Knudsen, M. 1915 *Ann. Phys., Lpz.,* **47**, 697.
Langmuir, J. 1916 *Phys. Rev.* **8**, 149.
Melville, H. W. & Gray, S. C. 1936*a* *Trans. Faraday Soc.* **32**, 271.
Melville, H. W. & Gray, S. C. 1936*b* *Trans. Faraday Soc.* **32**, 1026.
Miyamoto, S. 1933 *Trans. Faraday Soc.* **29**, 794.
Neumann, K. & Costeanu, V. 1939 *Z. phys. Chem.* A, **185**, 65.
Onsager, L. 1936 *J. Amer. Chem. Soc.* **58**, 1486.
Polanyi, M. & Wigner, E. P. (1928) *Z. phys. Chem.* A, **139**, 439.
Prüger, W. 1940 *Z. Phys.* **115**, 202.
Volmer, M. 1939 *Kinetik der Phasenbildung.* Leipzig: Steinkopf.

Dislocations as a cause of mechanical damping in metals

By J. D. Eshelby, *H. H. Wills Physics Laboratory, University of Bristol*

(*Communicated by N. F. Mott, F.R.S.—Received* 1 *December* 1948)

Zener has shown how thermoelastic effects give rise to damping of the mechanical vibrations of a solid. For example, in a vibrating reed opposite sides are alternately compressed and extended. This gives rise to an alternating temperature-difference across the width of the reed, and the resulting flow of heat leads to dissipation of mechanical energy.

In a vibrating single crystal of a metal an additional energy loss is observed which is usually attributed to the motion of dislocations. In the present paper the following mechanism is proposed. Dislocations are trapped in 'potential troughs' at the minima of the internal stress in the crystal. When the crystal vibrates the dislocations oscillate in their potential troughs and the moving stress-system associated with them produces a fluctuating temperature distribution in the material; this leads to damping as in Zener's case. The rate of loss of energy produced by a dislocation oscillating with given amplitude is calculated and the effect of a collection of them is discussed. An actual estimate of the damping in a vibrating crystal requires (i) a knowledge of the relation between the amplitude of oscillation of a dislocation and the vibrational stress causing it to move, and (ii) a knowledge of the density of dislocations in the material. A tentative discussion of (i) is given. The quantity (ii) is unknown; however, it is shown that the damping depends only on the ratio of the number of dislocations per unit area to the number of potential troughs per unit area. If this ratio is calculated from the theoretical result and the observed damping in copper single crystals, it is found to be of the order of unity. The present theory predicts that the damping should increase with frequency. This is in disagreement with the limited experimental data available.

Two subsidiary effects are also investigated, the thermoelastic damping arising from the interaction between the vibrational stresses and the stresses surrounding *stationary* dislocations, and the damping due to the emission of elastic waves from an oscillating dislocation. Both these effects are shown to be small compared with the thermoelastic damping caused by moving dislocations.

1. INTRODUCTION

A number of workers have measured the damping of mechanical vibrations in metal single crystals for low strain-amplitudes (about 10^{-7} to 10^{-5}) at various frequencies between 10 and 100 kc./sec. The damping is usually defined as the quantity Δ representing the energy dissipated per half-cycle divided by the total strain energy of the material. When the damping arising from large-scale thermal currents (Zener 1940) has been eliminated or allowed for, there still remains a loss which is generally supposed to be due to the motion of dislocations. For low amplitudes of vibration this loss is nearly independent of amplitude ('region of level decrement', in Read & Tyndall 1946), but increases at greater amplitudes. Only the low-amplitude damping is considered in this paper.

Attempts to explain this damping in terms of the inelastic strain (creep) observed in static tests (Seitz & Read 1941; Read & Tyndall 1946) have not been very successful; the observed values of Δ correspond to a much greater creep rate than is actually observed. The main object of this paper is to consider the contribution to the damping by 'elastically bound' dislocations, i.e. dislocations normally resting at the bottom of troughs in the potential distribution representing the internal stresses in the crystal lattice, and displaced from their equilibrium positions by an applied shear stress. Static stresses insufficient to lift a dislocation over the barrier separating two potential troughs, even with the aid of thermal fluctuations, will produce no permanent inelastic deformation. Alternating stresses of the same magnitude, arising from the vibration of the specimen in a damping test, will make the dislocation oscillate in its potential trough. This motion gives rise to energy loss by the mechanism discussed below, and thus a source of damping is provided which is not associated with large creep rates in static tests.

Several mechanisms can be imagined whereby oscillating dislocations could contribute to the dissipation of energy. The main one considered in this paper is a development of Zener's (1940) theory of thermoelastic damping. The movement of a dislocation alters the stresses at any point. This alteration of stress is associated with a change of temperature. The latter varies from point to point, so that temperature gradients exist in the material. The resulting flow of heat leads to dissipation of mechanical energy.

As a preliminary to the investigation of the mechanism of the last paragraph, the damping due to the interaction between the vibrational stress and the stresses surrounding *stationary* dislocations is investigated. Finally, the damping of a moving dislocation due to emission of elastic waves (analogous to radiation damping) is briefly considered.

The whole treatment is two-dimensional, i.e. the dislocations are all considered to be parallel and of infinite length.

2. THERMOELASTIC EFFECTS

Consider a material with zero thermal conductivity. It is shown in the Appendix (equations (a 1) and (a 3)) that the relation between the rate of change of temperature T at any point and the stresses there is

$$\frac{dT}{dt}=\frac{T}{c_p}\left\{\frac{1}{4}\left(\frac{\partial G^{-1}}{\partial T}\right)_p\frac{d(\sigma_0^2)}{dt}+\frac{9}{2}\left(\frac{\partial}{\partial T}\left(\frac{\nu}{E}\right)\right)_p\frac{d(p^2)}{dt}+\alpha\frac{dp}{dt}\right\}, \tag{1}$$

where p is the hydrostatic pressure (one-third of the sum of the diagonal terms of the stress tensor), σ_0^2 is the sum of the squares of the nine stress components, G, E and ν are the shear modulus, Young's modulus and Poisson's ratio, α is the volume expansion coefficient and T the absolute temperature. When the thermal conductivity is not zero the temperature can be found as a function of position and time by solving the heat-conduction equation with a distribution of heat sources of strength $c_p dT/dt$ per unit volume with dT/dt given by (1). If the resulting temperature varies from point to point heat currents will flow and the mechanical energy of the system will be dissipated.

If the material is initially unstressed a homogeneous stress varying harmonically with time will produce no loss, since there will be no temperature gradient. If the material contains an initial non-homogeneous stress there will be a non-homogeneous temperature distribution and hence energy loss, since p^2 and σ_0^2 will contain products of the initial and applied stress components.

This is the basis of the first damping mechanism considered below. The initial inhomogeneous stress is taken to be that of stationary dislocations, and the homogeneous stress is the stress due to the vibration of the specimen. The vibrational stress is not strictly homogeneous, but will be approximately so over the region in which its interaction with a dislocation is important.

The above mechanism postulates that the dislocations remain stationary in the material. If they move under the influence of the vibrational stresses the stress at any point will vary and give rise to additional temperature changes and consequently damping. This is the second mechanism considered.

3. Damping by stationary dislocations

Swift & Richardson (1947) have already suggested that the amplitude-independent damping which they observed for zinc crystals might arise from the interaction between the applied stress and the stresses surrounding the dislocations.

It is shown in the Appendix that a single dislocation in an infinite medium will, under the influence of an applied shear $\sigma_0 \sin \omega t$, produce an energy loss

$$w \sim \frac{a^2}{c_p T}\left(\frac{d \log G}{d \log T}\right)^2 \sigma_0^2 \tag{2}$$

per cycle per unit length. This holds apart from a numerical factor both for edge and screw dislocations.

Damping processes are usually discussed in terms of a set of relaxation times each associated with a certain oscillator strength. From this point of view the fact that w is independent of frequency has the following interpretation: the dislocation has a continuous relaxation spectrum, and the oscillator strength is the same for all relaxation times. Actually the spectrum will be cut off for relaxation times less l^2/κ, where l is the distance below which the elastic solution for the stresses surrounding the dislocation breaks down and κ is the thermal diffusivity of the material. Since $l \sim 10^{-7}$ cm. and $\kappa \sim 1$ cm.^{2}/sec. this could produce no observable effect at kilocycle frequencies.

The elastic energy per unit volume due to the vibrations is of the order of σ_0^2/G. If there are n dislocations threading unit area, and it is assumed that the contribution of each dislocation to the total loss is given by (2), then

$$\Delta \sim \frac{a^2 G}{c_p T}\left(\frac{d \log G}{d \log T}\right)^2 n. \tag{3}$$

With the following values for copper,

$$a = 3 \times 10^{-8}\,\text{cm.}, \qquad G = 4{\cdot}5 \times 10^{11}\,\text{dynes/cm.}^2,$$
$$c_p = 3{\cdot}4 \times 10^7\,\text{ergs/cm.}^3\,{}^\circ\text{C}, \qquad T = 300^\circ\,\text{K},$$
$$(d \log G/d \log T)^2 = 0{\cdot}02,$$

Δ is found to be of the order of $10^{-16}n$. The observed value of Δ, about 10^{-5}, would require about 10^{11} dislocations per cm.^{2}. Similar results hold for other materials.

Provided the mean distance between dislocations is appreciably greater than $(\kappa/\omega)^{\frac{1}{2}}$, the assumption that the total loss is the sum of the losses which each dislocation would produce in the absence of the others is justified. When they are closer together than this the calculation is rendered difficult for reasons discussed in the

Appendix. The correct value is not, however, likely to exceed (3). Since a density of 10^{11} dislocations per cm.^{2} seems excessive for an undistorted single crystal this mechanism will not be discussed further.

4. DAMPING BY OSCILLATING DISLOCATION

An expression is derived in the Appendix for the energy loss per cycle due to an oscillating dislocation. For a screw dislocation the result is

$$w = \frac{1}{2^7\pi^2}\frac{d^2\omega G a^2}{\kappa}\frac{a^2}{l^2}\frac{G}{c_p T}\left(\frac{d\log G}{d\log T}\right)^2 \quad \text{ergs per unit length.} \tag{3a}$$

The corresponding expression for the edge dislocation is rather complicated; it may be found from equation (a 24) in the Appendix. If the round value $\frac{1}{3}$ is taken for Poisson's ratio and it is assumed that

$$\frac{1}{G}\frac{dG}{dT} \doteqdot \frac{1}{E}\frac{dE}{dT} \gg \frac{1}{\nu}\frac{d\nu}{dT},$$

which is approximately correct for metals, the expression for the edge dislocation becomes, in ergs per unit length,

$$w \doteqdot \frac{1}{10}\frac{d^2\omega G a^2}{\kappa}\left\{\frac{1}{10}\frac{a^2}{l^2}\frac{G}{c_p T}\left(\frac{d\log G}{d\log T}\right)^2 + \frac{c_p - c_v}{c_p}\log\frac{\kappa}{\omega l^2}\right\}. \tag{4}$$

The symbols have the following meaning:

$\omega/2\pi$	frequency of oscillation.
d	amplitude of oscillation.
a	interatomic spacing.
G, E, ν	shear modulus, Young's modulus and Poisson's ratio.
c_p, c_v	specified heats per unit volume at constant pressure and volume respectively.
κ	thermal diffusivity (thermal conductivity divided by density and specific heat).
l	'cut-off' length.

The length l is a parameter introduced to allow for the fact that the expression derived from elasticity theory for the stresses surrounding a dislocation ceases to hold in its immediate neighbourhood. It represents approximately the distance from the dislocation below which the elastic solution ceases to hold good.

The main assumptions made in deriving these results are the following:

(i) The dislocation is of infinite length and oscillates with simple harmonic motion along its slip-plane. The medium in which it exists is of infinite extent, and there are no other dislocations present.

(ii) It carries with it the stress-system which surrounds it when stationary. This will be a good approximation if its maximum velocity is much less than that of sound in the material. That the velocity reached in damping experiments must be much less than that of sound can be seen as follows. The specimen used is about one-quarter wave-length long. A dislocation moving with the velocity of sound would therefore more than cover the whole length of the specimen in a half-cycle. Since the movement is presumably a very small fraction of the length of the specimen the velocity must be much smaller than that of sound.

(iii) The cut-off length l and amplitude of oscillation d are both much smaller than the length $(\kappa/\omega)^{\frac{1}{2}}$, about 10^{-3} cm., for metals in the range 10 to 100 kc./sec.

Approximate values of the constants occurring in (4) are, for copper at 300° K,

$$\frac{G}{c_p T} = 42, \quad \left(\frac{d \log G}{d \log T}\right)^2 = 0{\cdot}020, \quad \frac{c_p - c_v}{c_p} = 0{\cdot}027, \quad \kappa = 1{\cdot}2\,\text{cm.}^2\text{sec.}^{-1}.$$

With these values the { } of (4) becomes approximately

$$2{\cdot}7 \times 10^{-2}\left(3\frac{a^2}{l^2} + \log\frac{1{\cdot}2}{\omega l^2}\right).$$

The exact choice of l in the second term is not critical. For $l \sim 10^{-7}$ cm. and a frequency in the neighbourhood of 30 to 50 kc./sec. the logarithm will be of the order of 20. The elastic solution for the stresses surrounding the dislocation is generally supposed to hold to within a few atomic distances of its centre. 10^{-1} would be a good estimate for a^2/l^2; unity is certainly an over-estimate. It is probably justifiable to neglect the first term in comparison with the second, and this will be done in what follows. In the expression for the screw dislocation the logarithmic term is missing. It arises from the hydrostatic-pressure component of the stresses which is zero for a screw dislocation. The contribution from the screw dislocations will also be neglected.

According to the discussion in the Appendix the loss due to a number of dislocations situated at random in the material is approximately the sum of the losses they would produce separately. If the vibration shear stress in the material is σ_0 the elastic energy of the material per unit volume is of the order of $\frac{1}{2}\sigma_0^2/G$. If there are n dislocations threading unit area each giving the contribution (4) the damping will be

$$\Delta = \frac{1}{10\kappa}\frac{c_p - c_v}{c_p} G^2 a^2 \omega \log\left(\frac{\kappa}{\omega l^2}\right)\left(\frac{d}{\sigma_0}\right)^2 n. \tag{5}$$

If the amplitudes of the various dislocations differ, d^2 will represent an average value.

To find Δ explicitly it is necessary to know the value of d/σ_0. There is another phenomenon which should depend on this ratio, the reduction of shear modulus produced by the dislocations. A dislocation in moving a distance d across a unit cube produces a shear ad. If there are n dislocations threading unit area and they each move a distance d under the influence of a shear, the elastic strain σ_0/G will be accompanied by an additional strain nad, so that the apparent shear modulus G' will be given by

$$\frac{\sigma_0}{G'} = \frac{\sigma_0}{G} + nad.$$

The change in the inverse of the shear modulus on introducing the dislocations into a material previously containing none will be

$$\frac{\delta G^{-1}}{G^{-1}} = naG\frac{d}{\sigma_0}. \tag{5a}$$

Combined with (5) this gives

$$n = \frac{1}{10\Delta\kappa}\frac{c_p - c_v}{c_p}\,\omega\left(\frac{\delta G^{-1}}{G^{-1}}\right)^2 \log\frac{\kappa}{\omega l^2}. \tag{6}$$

This relation should be independent of any variation of d/σ_0 with frequency, provided G and Δ are measured at the same frequency. Since d^2 and d in (5) and (5a) represent averages, (6) should strictly be multiplied by $\overline{d^2}/\bar{d}^2$. There seem to be no measurements on single crystals which would serve to check these results.

Lawson (1941) investigated the change of damping and Young's modulus when compressive stresses of the order of 1 kg./mm.^{2} were applied to polycrystalline specimens of very pure copper. The crystal grains were several millimetres across, and Lawson considers that the effect of intercrystallite thermal currents can be neglected. The following data refer to the only case in which he gives numerical values both for damping and Young's modulus:

Δ before compression	2×10^{-3},
Δ after compression	4×10^{-3},
change of Young's modulus 6 to 7 %,	
frequency *c.* 50 kc./sec.	

It will be assumed that the proportional charges in G and Young's modulus are the same. The δG^{-1} in (6) refers to the total change when dislocations are introduced into a material originally free from them. If the comparison is to be made for the higher value of Δ, it is easy to see that the observed change of inverse modulus must be multiplied by 2. Inserted in (6) these figures give

$$n \doteqdot 5 \times 10^4 \text{ dislocations/cm.}^2.$$

Though there is no direct evidence for the number of dislocations in single crystals, this value is rather low compared with the value of about 10^6 generally assumed. Lawson's value of Δ is large compared to that found for copper single crystals (Reid 1941), about 10^{-4} to 10^{-5}. If some effect not present in single crystals is operative in his case the value 10^4 to 10^5 will be an underestimate.

According to a model which has been used successfully in discussing transient creep (Mott & Nabarro 1948) the dislocations are restrained from moving freely by stresses in the lattice. The variation of the stress with position is represented roughly by the expression

$$\sigma = \sigma_i \sin\frac{2\pi x}{\Lambda},$$

where σ_i is the maximum internal strain and Λ is the distance between successive maxima. The dislocations occupy the points where $\sigma = 0$. When an external shear stress σ_0 is applied these points move a distance $d = \Lambda\sigma_0/2\pi\sigma_i$. If the dislocation is situated in a unit cube it produces a shear ad in moving the distance d. The effective force on the dislocation is given by $Fd = \sigma_0 ad$ or $F = \sigma_0 a$, and the dislocation can be regarded as restrained by an elastic restoring force $2\pi\sigma_i a/\Lambda$ per unit displacement per unit length, or as occupying the bottom of a trough in a potential field

$$V = (\sigma_i a\Lambda/2\pi)\cos 2\pi x/\Lambda. \tag{7}$$

For slow movement of the dislocation d/σ_0 should thus have the value $\Lambda/2\pi\sigma_i$. However, the inertia of the material set in motion by the movement of the dislocation cannot in general be neglected. A dislocation can be shown to have an equivalent mass of the order of ρa^2 per unit length, where ρ is the density of the material. Consider the normal equation of forced harmonic motion

$$m\ddot{x}+r\dot{x}+kx = F\sin\omega t. \tag{8}$$

The energy dissipated per cycle is $2\pi r d^2\omega$, where d is the amplitude of the oscillation. This is just of the form (4) apart from the logarithmic term which varies only slowly with ω. It is reasonable to suppose that the motion of the dislocation is governed by (8) with

$$m = \rho a^2, \quad r = \beta G a^2/2\pi, \quad k = 2\pi\sigma_i a/\Lambda,$$

where

$$\beta = \frac{1}{10\kappa}\frac{c_p-c_v}{c_p}\log\frac{\kappa}{\omega l^2}.$$

The natural undamped frequency of the motion is given by

$$\omega_0 = (k/m)^{\frac{1}{2}} = (2\pi\sigma_i/\rho\Lambda a)^{\frac{1}{2}}.$$

At very low frequencies $(d/\sigma_0)^2 = (\Lambda/2\pi\sigma_i)^2$. At other frequencies this value must be multiplied by the factor

$$\{(1-\omega^2/\omega_0^2)^2+r^2\omega^2/k^2\}^{-1}. \tag{9}$$

In terms of the quantity

$$z = G\Lambda/2\pi\sigma_i,$$

the two parameters in (9) have the values

$$\omega_0 = (G/\rho a)^{\frac{1}{2}} z^{-\frac{1}{2}} \sim 10^{10} z^{-\frac{1}{2}}, \quad r/k = (\beta a/2\pi)\, z \sim 10^{-9} z. \tag{10}$$

According to the model being used σ_i represents the critical shear stress of the crystal. G/σ_i usually lies between 10^2 and 10^5 for single crystals. Presumably Λ lies between 10^{-2} and 10^{-6} cm. Then z will lie between 10^3 and 10^{-4}. Thus for frequencies in the kilocycle range ω/ω_0 and $r\omega/k$ will always be small compared with unity and the low-frequency value $d/\sigma_0 = \Lambda/2\pi\sigma_i$ may be used. Equation (5) therefore becomes

$$\Delta = \frac{\omega a^2}{10\kappa}\frac{c_p-c_v}{c_p}\log\frac{\kappa}{\omega l^2}\left(\frac{G\Lambda}{2\pi\sigma_i}\right)^2 n. \tag{11}$$

Of the quantities in (11) σ_i can be found from the yield strength and the exact value of l is unimportant. The remainder are known constants except for Λ and n which are quite unknown. The combination $n' = \Lambda^2 n$, however, represents the proportion of potential troughs occupied by dislocations. The value of n' calculated from

$$n' = \Delta\left[\frac{\omega a^2}{10\kappa}\frac{c_p-c_v}{c_p}\log\frac{\kappa}{\omega l^2}\left(\frac{G}{2\pi\sigma_i}\right)^2\right]^{-1} \tag{12}$$

should give some idea of the reasonableness of the assumptions on which (11) is based.

For a single crystal of copper at 33·5 kc./sec. Read (1941) found $\Delta \sim 10^{-4}$ for the freshly mounted specimen, falling to $\Delta \sim 10^{-5}$ on annealing. Seitz (1943) gives the yield stress 0·1 kg./mm.^{2} for a copper crystal of comparable purity. Using these figures, together with the values already quoted, for the remaining constants gives

$$n' = 2 \times 10^3 \Delta,$$

or $\frac{1}{5}$ for the unannealed and $\frac{1}{50}$ for the annealed specimen. A result of the order of one dislocation in each potential trough in the unannealed material seems reasonable. A value greatly in excess of unity is unlikely, since the dislocations would then be substantially controlled by each other's stresses so that effectively $\sigma_i \sim Ga/\Lambda$ and $n \sim \Lambda^{-2}$; this is easily seen to lead to $\Delta \propto n^{-1}$, or a decrease of damping with number of dislocations. Read found, on the contrary, that Δ increased on subjecting his specimens to static stresses and which presumably increased the number of dislocations.

A large number of measurements have been made on zinc (Read 1941; Read & Tyndall 1946; Swift & Richardson 1947). This material shows strong 'self-annealing' even at room temperature; for example, a specimen had a Δ of 10^{-3} on first mounting in the apparatus, 7×10^{-5} after 30 hr. and 6×10^{-6} after 7 days. Because of its great elastic and thermal anisotropy and the fact that the stresses surrounding a dislocation in a hexagonal crystal have not been calculated, it would be rash to assume that the present theory can be applied. However, it may be of interest to compare the values of n' obtained from (12) using bulk values of the constants involved. Read & Tyndall found a departure from Hooke's law at a shear strain of about 8×10^{-6}; Swift & Richardson adopt the same value in discussing their results. If this strain is identified with σ_i/G, (12) gives $n' = 45\Delta$. The observed Δ varies between about 3×10^{-3} and 3×10^{-6}, giving an n' between 10^{-1} and 10^{-4}.

According to (11) Δ is approximately proportional to frequency. No account has been taken of this when comparing theory and experiment in the last few paragraphs, since most measurements have in fact been made at isolated frequencies.

Read (1941) has published curves for a zinc crystal at 39 and 78 kc./sec., showing that Δ is inversely proportional to the frequency. In view of the great differences between specimen and specimen, both in the value of Δ and its variation with amplitude, this relation cannot be assumed to hold universally. On the other hand, according to Seitz (1943) Δ is independent of frequency for copper; there seems to be no reference to this in the literature. Clearly more experimental results are needed, but it may be of interest to consider what modifications (apart from dropping the limitation to a two-dimensional problem) would be needed to bring the present theory into line with these results if confirmed.

The frequency dependence in (11) arises essentially from two facts. First, the thermoelastic effects are equivalent to a resistive force on the dislocation approximately proportional to its velocity; this is confirmed by calculations for a dislocation moving with uniform velocity. Secondly, the natural frequency of oscillation of a dislocation in its potential trough is much higher than the applied frequency.

If ω_0 could be taken to be of the order of the frequency used in damping measurements Δ would no longer be proportional to frequency, being given by the product of (11) and the expression (9). Some rather indirect evidence for such a low value of ω_0 is provided by a comparison of observed creep rates with theory. The unit process in slip is considered to be the movement of a dislocation from trough to trough of the potential (7) under the influence of thermal fluctuations and an applied stress. The variation of creep rate with temperature might therefore be expected to be given by a rate-process formula of the type $f\exp(-U/kT)$. If experimental data are fitted to such an expression (Kauzmann 1941) the frequency factor f is of an altogether lower order of magnitude than the ω_0 estimated from (10).

5. RADIATION DAMPING

An isolated screw dislocation oscillating in an infinite medium dissipates energy at a rate

$$w = \pi\rho a^2 d^2 \omega^2/16$$

per cycle per unit length as a result of the radiation of elastic waves (Appendix, § A 6). ρ is the density of the medium, and the other symbols are as in § 4. Compared with the thermoelastic dissipation (3 a) this is extremely small at kilocycle frequencies. The radiation damping for an edge dislocation will be of the same order and hence will be small in comparison with the thermoelastic damping (4). In actual damping experiments the wave-length of the emitted elastic waves is greater than the dimensions of the specimen, and the calculation for the infinite specimen does not apply. Reflexion at the surface of the specimen would not produce energy loss but a reactive effect on the motion of the dislocation. Scattering by lattice imperfections (e.g. the centres of other dislocations where the stress-strain relation is non-linear) might ultimately lead to dissipation.

6. CONCLUSION

The thermoelastic effect arising from stationary dislocations is inadequate to explain the damping observed in single crystals.

The corresponding effect for moving dislocations seems capable of giving an effect of the right order of magnitude, but the frequency dependence needs more detailed investigation both theoretically and experimentally.

The experimental methods hitherto used do not lend themselves to measurements of damping and elastic constants over a range of frequencies on the same specimen without removal from the apparatus. A technique which could give such results might throw valuable light on the dynamics of moving dislocations.

APPENDIX

A 1. *Method of obtaining energy loss*

The first step is the calculation of the temperature changes associated with the varying elastic conditions in the medium. The state of the medium is determined by the temperature T and the stress tensor p_{ij}. For the entropy change of unit volume we have

$$dS = \left(\frac{\partial S}{\partial T}\right)_p dT + \left(\frac{\partial S}{\partial p_{ij}}\right)_T dp_{ij}.$$

The convention that a repeated index is summed for the values 1, 2, 3 will be used. The second term on the right therefore represents nine terms and in the calculations below

$$(p_{ij})^2 = p_{ij}p_{ij} = p_{11}^2+\dots+2p_{12}^2+\dots$$

and

$$p_{mm} = p_{11}+p_{22}+p_{33}.$$

The coefficient of dT is by definition c_p/T. The coefficient of dp_{ij} can be transformed by considering the Gibbs function, the change of which is given by

$$d\phi = -e_{ij}dp_{ij}-S\,dT.$$

Here e_{ij} is the strain tensor defined in terms of the displacement vector u_i by the relation

$$e_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right).$$

Since $d\phi$ is a perfect differential

$$\left(\frac{\partial e_{ij}}{\partial T}\right)_p = \left(\frac{\partial S}{\partial p_{ij}}\right)_T.$$

e_{ij} and p_{ij} are related by

$$e_{ij} = \epsilon p_{ij}+\zeta p_{mm}\delta_{ij}+\tfrac{1}{3}\alpha(T-T_0)\,\delta_{ij},$$

where

$$\epsilon = \tfrac{1}{2}G, \quad \zeta = \nu/E. \tag{a 1}$$

G and E are the shear modulus and Young's modulus and ν is Poisson's ratio. The final term represents the thermal expansion; α is the coefficient of volume expansion and T_0 some arbitrary temperature at which the material is considered to be unstrained in the absence of stress. Hence

$$dS = \frac{c_p}{T}dT+\{\epsilon' p_{ij}+\zeta' p_{mm}\delta_{ij}+\tfrac{1}{3}\alpha\delta_{ij}\}\,dp_{ij}, \tag{a 2}$$

where

$$\epsilon' = \frac{1}{2}\frac{dG^{-1}}{dT}, \quad \zeta' = \frac{d}{dt}\left(\frac{\nu}{E}\right).$$

If the second term in (a 2) is integrated from an initial state p_{ij}^0 to a final state p_{ij} and multiplied by T/c_p, we obtain a quantity

$$v_0 = \frac{T}{c_p}\{\tfrac{1}{2}\epsilon'(p_{ij}^2-p_{ij}^{02})+\tfrac{1}{2}\zeta'[(p_{mm})^2-(p_{mm}^0)^2]+\tfrac{1}{3}\alpha(p_{mm}-p_{mm}^0)\}, \tag{a 3}$$

and the entropy equation can be written in the form

$$T\,dS = c_p(dv+dv_0),$$

where dv has been written for dT. v will be used to denote the change of temperature from the mean.

If the thermal conductivity is k then $k\nabla^2 v$ is the rate of flow of heat into unit volume. If we identify this with $T\,dS$ the last equation can be written as

$$\kappa\nabla^2 v-\dot{v}+\dot{v}_0 = 0, \tag{a 4}$$

where $\kappa = k/c_p$ is the thermal diffusivity. Clearly v_0 represents the value of v when the conductivity is zero. When v is known the rate of dissipation of energy can be found from the increase of entropy. If T is the mean absolute temperature the rate of increase of entropy per unit volume is

$$\frac{k^2\nabla^2 v}{T+v} = \frac{k\nabla^2 v}{T} - \frac{1}{T^2} kv\nabla^2 v,$$

if higher terms are neglected. The total change of entropy is found by integration over the whole volume. Application of Green's theorem converts the first term into an integral over the boundary of the solid. If $\operatorname{grad} v = 0$ over the boundary this term gives no contribution to the energy loss. We shall assume that this is the case. The rate of dissipation of energy is thus

$$-\frac{k}{T}\int_{\mathbf{r}} v\nabla^2 v\, d\mathbf{r} = \frac{c_p}{T}\int_{\mathbf{r}} v(\dot{v}_0 - \dot{v})\, d\mathbf{r}$$

by (a 4). For a cyclic process with period $2\pi/\omega$ the energy loss per cycle is

$$w = \frac{c_p}{T}\int_0^{2\pi/\omega}\int_{\mathbf{r}} v(\dot{v}_0 - \dot{v})\, d\mathbf{r}\, dt = \frac{c_p}{T}\int_0^{2\pi/\omega}\int_{\mathbf{r}} v\dot{v}_0\, d\mathbf{r}\, dt. \qquad \text{(a 5)}$$

A more rigorous discussion is given by Päsler (1944).

For a two-dimensional problem in which the medium is of infinite extent w is most easily calculated by making a two-dimensional Fourier analysis of v and v_0. w can then be found without calculating v explicitly. For any function of position $f(\mathbf{r})$ we define the Fourier transform $\bar{f}(\mathbf{k})$ by

$$\bar{f}(\mathbf{k}) = \frac{1}{2\pi}\int_{\mathbf{r}} f(\mathbf{r})\, e^{i\mathbf{k}.\mathbf{r}}\, d\mathbf{r}, \qquad \text{(a 6)}$$

the integral extending over all space. If f depends on the time t as a parameter so will $\bar{f}$. We need two results from the theory of Fourier transforms (cf. Titchmarsh 1937). The first is the Fourier integral theorem which states that

$$f(\mathbf{r}) = \frac{1}{2\pi}\int_{\mathbf{k}} \bar{f}(\mathbf{k})\, e^{-i\mathbf{k}.\mathbf{r}}\, d\mathbf{k} = \frac{1}{(2\pi)^2}\int_{\mathbf{k}} d\mathbf{k}\int_{\mathbf{r}} f(\mathbf{r}')\, e^{i\mathbf{k}.(\mathbf{r}'-\mathbf{r})}\, d\mathbf{r}', \qquad \text{(a 7)}$$

and the second is Parseval's theorem

$$\int_{\mathbf{k}} f_1(\mathbf{r}) f_2^*(\mathbf{r})\, d\mathbf{r} = \int_{\mathbf{k}} \bar{f}_1(\mathbf{k})\, \bar{f}_2^*(\mathbf{k})\, d\mathbf{k}, \qquad \text{(a 8)}$$

where * denotes the complex conjugate.

Since $\nabla^2 e^{i\mathbf{k}.\mathbf{r}} = -k^2 e^{i\mathbf{k}.\mathbf{r}}$, (a 4) in conjunction with (a 7) gives

$$\kappa k^2 \bar{v} + \dot{\bar{v}} = \dot{\bar{v}}_0. \qquad \text{(a 9)}$$

As v and v_0 are real (a 8) applied to (a 5) gives

$$w = \frac{c_p}{T}\int_0^{2\pi/\omega}\int_{\mathbf{k}} \bar{v}\dot{\bar{v}}_0^*\, d\mathbf{k} = \frac{c_p\kappa}{T}\int_0^{2\pi/\omega}\int_{\mathbf{k}} |\bar{v}|^2 k^2\, d\mathbf{k}; \qquad \text{(a 10)}$$

the second equation follows from (a 9).

The following result will also be needed. If the function $f(\mathbf{r})$ is shifted as a whole by a vector displacement $\mathbf{d}$, becoming $f(\mathbf{r}-\mathbf{d})$, then the new transform is given by

$$\begin{aligned} \bar{f}'(\mathbf{k}) &= \frac{1}{2\pi}\int_{\mathbf{r}} f(\mathbf{r}-\mathbf{d})\,\mathrm{e}^{i\mathbf{k}.\mathbf{r}}\,d\mathbf{r} \\ &= \frac{1}{2\pi}\int_{\mathbf{r}'} f(\mathbf{r}')\,\mathrm{e}^{i\mathbf{k}.(\mathbf{r}'+\mathbf{d})}\,d\mathbf{r}' \\ &= \mathrm{e}^{i\mathbf{k}.\mathbf{d}}\bar{f}(\mathbf{k}). \end{aligned} \qquad \text{(a 11)}$$

A 2. *Stresses around edge and screw dislocations*

For convenience the stresses round a dislocation are given below, together with the combinations of them which will be required later.

(i) *Edge dislocation.* The components in the xy plane have been given by Koehler (1941). The remaining components have been calculated on the assumption of plane strain:

$$\left.\begin{aligned} p_{11} &= -D(\sin 3\theta + 3\sin\theta)/2r, & p_{33} &= -D(4\nu\sin\theta)/2r, \\ p_{22} &= D(\sin 3\theta - \sin\theta)/2r, & p_{13} &= p_{23} = 0, \\ p_{12} &= D(\cos 3\theta + \cos\theta)/2r, & p_{mm} &= -D\{4(1+\nu)\sin\theta\}/2r, \\ & p_{ij}^2 = 2D^2\{(1+\nu^2)-\nu^2\cos 2\theta\}/r^2, \end{aligned}\right\} \qquad \text{(a 12)}$$

where $D = Ga/2\pi(1-\nu)$ and a is the interatomic spacing.

(ii) *Screw dislocation.* The displacement has the simple form (Burgers 1939)

$$u_3 = \frac{a}{2\pi}\arctan\frac{y}{x} = \frac{a}{2\pi}\theta, \quad u_1 = 0, \quad u_2 = 0. \qquad \text{(a 13)}$$

The stresses can easily be derived; all vanish but two:

$$\left.\begin{aligned} p_{23} &= \frac{aG}{2\pi}\frac{\cos\theta}{r}, & p_{13} &= \frac{aG}{2\pi}\frac{\sin\theta}{r}, \\ p_{mm} &= 0, & p_{ij}^2 &= \frac{a^2G^2}{2\pi^2r^2}. \end{aligned}\right\} \qquad \text{(a 14)}$$

A 3. *Energy loss due to stationary dislocations*

If a material has a large initial strain p_{ij}^0 and a small additional strain p_{ij}' is applied’ (a 3) gives

$$v_0 = \frac{T}{c_p}\{\tfrac{1}{2}\epsilon' p_{ij}^0 p_{ij}' + \tfrac{1}{2}\zeta'(p_{mm}^0)(p_{mm}') + \tfrac{1}{3}\alpha(p_{mm}')\}.$$

In our application p_{ij}^0 represents the stresses surrounding a dislocation and p_{ij}' is the stress arising from the vibration of the material. As the simplest case consider a screw dislocation subject to the alternating shear $p_{23}' = p_{32}' = S\cos\omega t$. Then by (a 14)

$$v_0 = A\frac{\cos\theta}{r}\cos\omega t, \quad A = \frac{\epsilon' aGST}{2\pi c_p}.$$

From (a 6)

$$\bar{v}_0 = \frac{A}{2\pi}\cos\omega t\int_0^\infty dr\int_0^{2\pi}\mathrm{e}^{ikr}\cos\theta\,d\theta.$$

If we introduce polar co-ordinates (k, θ') in **k**-space and $\phi = \theta - \theta'$,

$$\bar{v}_0 = \frac{A}{2\pi}\cos\omega t \int_0^\infty dr \int_0^{2\pi} \mathrm{e}^{ikr\cos\phi}\cos(\theta' + \phi)\, d\phi$$

$$= A\cos\omega t\cos\theta' \int_0^\infty J_1(kr)\, dr = A\,\frac{\cos\theta'}{k}\cos\omega t.$$

The steady-state solution of (a 9) is

$$\bar{v} = \frac{A\omega\cos\theta'}{k}\frac{\cos(\omega t + \delta)}{(\kappa^2 k^4 + \omega^2)^{\frac{1}{2}}}, \quad \tan\delta = \frac{\omega}{\kappa k^2}.$$

Hence from (a 10)

$$w = \frac{A^2\omega^2 c_p \kappa}{T}\int_0^{2\pi/\omega}\cos^2\omega t\, dt \int_0^{2\pi}\cos^2\theta'\, d\theta' \int_0^\infty \frac{k\, dk}{\kappa^2 k^4 + \omega^2} = \frac{\pi^3 A^2 c_p}{4T}. \qquad \text{(a 15)}$$

It is easy to take account of the fact that v_0 does not in fact increase as $1/r$ at small distances from the origin. For example, if $1/r$ is replaced by $(1 - \mathrm{e}^{-r/l})/r$, the expression for $\bar{v}_0$ becomes multiplied by $(1 + k^2 l^2)^{-\frac{1}{2}}$. w is then equal to (a 15) multiplied by the factor

$$(1 + 4x^2\pi^{-1}\log x)(1 + x^4)^{-1}, \quad \text{where} \quad x = (l^2\omega/\kappa)^{\frac{1}{2}}.$$

For frequencies appreciably less than κ/l^2 the expression for w is practically unaltered. Since for a metal $\kappa \sim 1$ and l (the distance from the centre of the dislocation within which the elastic solution breaks down) is certainly less than 10^{-6} cm., no error will be made by taking $l = 0$ for kilocycle frequencies. Essentially the same result follows if $1/r$ is replaced by any function which remains finite at the origin and approaches $1/r$ for $r > l$.

In full (a 15) reads

$$w = \frac{\pi}{64}\left(\frac{d\log G}{d\log T}\right)^2 \frac{a^2 S^2}{T c_p}. \qquad \text{(a 16)}$$

Other cases in which an edge or screw dislocation is subjected to various kinds of shear can be treated similarly. For example, if an edge dislocation is subjected to a shear of magnitude S and any direction in the xy plane so that

$$p'_{12} = S\cos\omega t\sin 2\phi, \quad p'_{11} = -p'_{22} = S\cos\omega t\cos 2\phi,$$

where ϕ is an arbitrary angle, then the loss per cycle is obtained from (a 16) by replacing $\pi/64$ by $\pi/32(1-\sigma)^2$. The results are of the same order of magnitude in all cases. If the diagonal terms p_{mm} are involved the factor $d\log(E/\sigma)/d\log T$ will occur; it is of the same order of magnitude as $d\log G/d\log T$. The actual value of v can be found with the aid of (a 7). The result is

$$v = \frac{A\cos\theta}{r}\{(1 + x\,\mathrm{ker}'\,x)\cos\omega t - x\,\mathrm{kei}'\,x\sin\omega t\}, \quad x^2 = \frac{r^2\omega}{\kappa}.$$

From the properties of the ker and kei functions it is easy to show that as $\kappa \to 0$, $v \to v_0$ and as $\kappa \to \infty$, $v \to 0$ as might be expected. As $\mathrm{ker}\,x$ and $\mathrm{kei}\,x$ fall off exponentially for $x > 1$, it follows that $(v_0 - v)$ is only appreciable within a cylinder of radius $\sim (\kappa/\omega)^{\frac{1}{2}}$ about the axis of the dislocation.

A 4. *Energy loss due to a moving dislocation*

Consider a dislocation moving along the X-axis as slip plane and carrying with it the stress-system it has when stationary. Let the position of its centre be $x = d\sin\omega t$. Then by (a 11)

$$\bar{v}_0(t) = \bar{v}_0(0)\exp\{ikd\sin\omega t\cos\theta'\}. \tag{a 17}$$

For a periodic (but not necessarily sinusoidal) $\bar{v}_0$ the general periodic solution of (a 9) is

$$\bar{v}(t) = \int_{-\infty}^{t} \dot{\bar{v}}_0(\tau)\exp\{-\kappa k^2(t-\tau)\}\,d\tau.$$

From (a 17)

$$\dot{\bar{v}}_0(\tau) = ikd\omega\cos\omega\tau\cos\theta'\bar{v}_0(0)\exp\{ikd\sin\omega\tau\cos\theta'\},$$

and hence

$$\overline{v\dot{v}_0^*} = d^2\omega^2k^2\cos\omega t\cos^2\theta'\,|\,\bar{v}_0(0)\,|^2\int_{-\infty}^{t}\cos\omega\tau\exp\{-\kappa k^2(t-\tau)\}(\cos kq+i\sin kq)\,d\tau, \tag{a 18}$$

where

$$q = d(\sin\omega\tau-\sin\omega t)\cos\theta', \tag{a 19}$$

from which the energy loss per cycle can be found by (a 10).

For an edge dislocation (a 12) gives, neglecting the effect of any applied forces,

$$v_0(0) = Ar^{-2}+Br^{-2}\cos 2\theta+Cr^{-1}\sin\theta, \tag{a 20}$$

where

$$A = D^2\{(1+\sigma^2)\epsilon'+(1+\sigma)^2\zeta'\}\,T/c_p,$$
$$B = -D^2\{\epsilon'\sigma^2+\zeta'(1+\sigma)^2\}\,T/c_p,$$
$$C = \tfrac{2}{3}\alpha D(1+\sigma)\,T/c_p.$$

As the simplest way of taking into account the fact that the stresses are finite at the origin we replace r by $\rho = (r^2+l^2)^{\frac{1}{2}}$. Then (a 6) gives

$$\begin{aligned}\bar{v}_0(0) &= \frac{1}{2\pi}\int_0^\infty r\,dr\int_0^{2\pi}\exp\{ikr\cos\phi\}\,d\phi\,\{A\rho^{-2}+B\rho^{-2}\cos 2(\theta'+\phi)+C\rho^{-1}\sin(\theta'+\phi)\}\\ &= \int_0^\infty r\,dr\{A\rho^{-2}J_0(kr)-B\rho^{-2}J_2(kr)\cos 2\theta'-iC\rho^{-1}J_1(kr)\sin\theta'\}\\ &= AH_1-BH_2\cos 2\theta'-iCH_3\sin\theta',\end{aligned}$$

where

$$H_1 = \int_0^\infty \frac{J_0(kr)}{r^2+l^2}\,r\,dr = K_0(kl)$$

by W. B. F., p. 425, eqn. (5),†

$$H_2 = \int_0^\infty \frac{J_2(kr)}{r^2+l^2}\,r\,dr = 2(kl)^{-2}-K_2(kl),$$

by a modification of the method of W. B. F., § 13·51, taking into account the singularity at the origin, and

$$H_3 = \int_0^\infty \frac{J_1(kr)}{(r^2+l^2)^{\frac{1}{2}}}\,r\,dr = \tfrac{1}{2}l\{I_0K_1-I_1K_0\},$$

---

† Here and elsewhere W. B. F. refers to the book by Watson (1922).

where the argument of the I's and K's is $\frac{1}{2}kl$. This last result is most easily found from W. B. F., p. 435, eqn. (3), with $\nu = 0$ on differentiating with respect to a.

The integrations required by (a 10) will now be carried out on the assumption that

$$d^2\omega/\kappa \leqslant 1 \tag{a 21}$$

and

$$l^2\omega/\kappa \leqslant 1, \tag{a 22}$$

but with no assumption about the relative size of d and l. We have

$$|\bar{v}_0(0)|^2 = A^2H_1^2 - 2ABH_1H_2\cos 2\theta' + B^2H_2^2\cos^2 2\theta' + C^2H_3^2\sin^2\theta'.$$

In (a 18) this factor and $\cos^2\theta'$ are even functions of θ', whilst the term $i\sin kq$ is odd, so that this term will not contribute to the θ'-integration and can be omitted. Consider the ratio of the coefficients of k in the factors $\cos kq$ and $\exp\{-\kappa k^2(t-\tau)\}$. It is

$$x = d(\sin\omega\tau - \sin\omega t)\cos\theta'/\kappa^{\frac{1}{2}}(t-\tau)^{\frac{1}{2}}.$$

The values of τ and t which occur satisfy

$$-\infty < \tau \leqslant t \leqslant 2\pi/\omega.$$

For the critical case $\tau = t$, $x = 0$. In other cases x is of the order of $d(\omega/\kappa)^{\frac{1}{2}}$. If (a 21) holds the argument of the cosine is always much smaller than that of the exponential. When $\bar{v}\bar{v}_0^*$ is integrated with respect to k the contributions from k-values for which $\kappa k^2(t-\tau) \geqslant 1$ will be small. Hence in the k-integration $\cos kq$ can be replaced by unity. Thus

$$\begin{aligned} w &= \frac{c_p d^2\omega^2}{T}\int_0^{2\pi/\omega}\exp\{-\kappa k^2 t\}\cos\omega t\,dt\int_0^{2\pi}\cos^2\theta'\,d\theta'\int_0^\infty|\bar{v}_0(0)|^2 k^3\,dk \\ &\qquad\qquad\times\int_{-\infty}^{t}\exp\{\kappa k^2\tau\}\cos\omega\tau\,d\tau \\ &= \frac{\pi c_p d^2\omega\kappa}{T}\int_0^{2\pi}\cos^2\theta'\,d\theta'\int_0^\infty|\bar{v}_0(0)|^2\frac{k^5\,dk}{\kappa^2k^4+\omega^2} \\ &= \frac{\pi^2 c_p d^2\omega\kappa}{T}\int_0^\infty\{A^2H_1^2 - ABH_1H_2 + \tfrac{1}{2}B^2H_2^2 + \tfrac{1}{4}C^2H_3^2\}\frac{k^5\,dk}{\kappa^2k^4+\omega^2} \\ &= \frac{\pi^2 c_p d^2\omega}{Tl^2\kappa}\{A^2L_1 - ABL_2 + \tfrac{1}{2}B^2L_3 + \tfrac{1}{4}C^2L_4\}, \end{aligned} \tag{a 23}$$

say. If $X = (x^4+\xi^4)^{-1}$ with $\xi = (l^2\omega/\kappa)^{\frac{1}{2}}$ then

$$L_1 = \int_0^\infty xK_0^2(x)\,x^4X\,dx = \int_0^\infty xK_0^2(x)\,dx - \xi^4\int_0^\infty xK_0^2(x)\,X\,dx.$$

The first integral has the value $\frac{1}{2}$. The second must be less than

$$\xi^4\int_0^\infty X\,dx = \pi\xi/\sqrt{2},$$

since $0 \leqslant xK_0(x) < 1$ for $0 \leqslant x$. Thus

$$L_1 = \tfrac{1}{2} + O(\xi).$$

Similarly
$$L_2 = \int_0^\infty xK_0(x)\{2x^{-2} - K_2(x)\}\,dx + O(\xi) = \frac{\pi^2}{12} - \tfrac{1}{2} + O(\xi)$$

and
$$L_3 = \int_0^\infty x\{K_2(x) - 2x^{-2}\}^2\,dx + O(\xi) = \tfrac{1}{2} + O(\xi).$$

The relation
$$I_0(x)\,K_1(x) + I_1(x)\,K_0(x) = x^{-1}$$

gives
$$L_4/l^2 = \int_0^\infty x\{x^{-2} - 4I_0K_0I_1K_1\}\,x^4Y\,dx,$$

where $Y = (x^4 + \xi^4/16)^{-1}$. This can be split up to give

$$L_4/l^2 = \int_0^1 x^3Y\,dx - 4\int_0^1 xI_0K_0I_1K_1\,Y\,dx + \int_1^\infty x\{x^{-2} - 4I_0K_0I_1K_1\}\,dx$$
$$+\tfrac{1}{4}\xi^4\int_0^1 xI_0K_0I_1K_1\,Y\,dx - \tfrac{1}{4}\xi^4\int_1^\infty x\{x^{-2} - 4I_0K_0I_1K_1\}\,Y\,dx.$$

The first term is a standard form, the second and third can be evaluated numerically and the last two can be shown to be of the order of ξ. In this way it is found that

$$L_4/l^2 = -\log\xi + 0{\cdot}109 + O(\xi).$$

Hence if (a 22) holds

$$w = \frac{\pi^2 c_p d^2\omega}{2\kappa T}\left\{l^{-2}(A^2 - 0{\cdot}645AB + \tfrac{1}{2}B^2) + \tfrac{1}{4}C^2\left(\log\frac{\kappa}{\omega l^2} + 0{\cdot}218\right)\right\}. \qquad \text{(a 24)}$$

For a screw dislocation (a 14) gives

$$v_0(0) = Ar^{-1}, \quad \text{where now} \quad A = T\epsilon' a^2G^2/4\pi^2c_p.$$

The loss per cycle can be found from (a 24) by giving A this value and setting $B = C = 0$.

Other choices of the 'cut-off' function ρ lead to similar results. For the term in C^2 if

$$\rho^{-1} = r^{-1}\{1 - \exp(-r/l)\} \quad \text{or} \quad \rho^{-1} = \begin{cases} r^{-1} & (r \geqslant l), \\ 0 & (r < l), \end{cases}$$

it can be shown that the quality $0{\cdot}218$ in (a 24) is replaced by zero or $0{\cdot}232$ respectively. If $\log(\kappa/\omega l^2)$ is of the order of 10 the loss arising from the C^2 term is very insensitive to the choice of the form of ρ or the value of l. The terms involving A and B are naturally more sensitive, since l occurs in the form l^{-2}. It is also difficult to find analytical expressions for ρ which are easy to deal with. A rather artificial choice is

$$\rho^{-2} = r^{-2}\{1 - J_0(r/l)\},$$

which is finite at the origin and approaches the value r^{-2} when r is a few multiples of l. Using W.B.F., p. 406, eqn. (11), it is easy to show that in the A^2 term the effect of this choise of ρ is to replace the l of (a 24) by $l\sqrt{2}$. The terms in A and B will be found to be smaller than the C^2 term, so that more detailed calculations are unnecessary.

A 5. *The effect of a large number of dislocations*

If an infinite block of material contains a large number of dislocations, the temperature variation is given by the sum of the v's of the individual dislocations. Cross-terms between these will occur when w is found from (a 5) or (a 10).

The method used so far can be summed up as follows. Instead of dealing with the temperature variation v directly we construct the transformed function $\bar{v}$ in reciprocal **k**-space. From this w is derived by integration according to (a 10). This is analogous to the method used in X-ray analysis, where the transform of the electron density of the scattering material is plotted in **k**-space and the distribution of scattered radiation is determined by calculation in **k**-space. The problem of a more or less random distribution of dislocations is thus analogous to the problem of X-ray scattering from an amorphous solid. The calculation below of the damping produced by a large number of dislocations is based on the averaging process used in the X-ray case (cf. Debye 1927).

In the case of stationary dislocations producing damping by interaction with an applied stress there is a fundamental difficulty. If a dislocation is reversed in sign so is its temperature variation. A macroscopically undistorted material will contain about equal numbers of dislocations of either sign. This makes any calculation of the total damping difficult. Since the effect of stationary dislocations seems to be unimportant this case will not be pursued. This much can be said, however. It has been seen that the expression $(v-v_0)$ extends only to a distance of the order of $(\kappa/\omega)^{\frac{1}{2}}$ about each dislocation; provided, therefore, the mean distance between dislocations is much greater than this distance the loss will be effectively the sum of the losses produced by each dislocation in the absence of the others.

We next consider a set of edge dislocations oscillating under the influence of an applied alternating shear. For simplicity we assume that they have a choice of two slip-planes at right angles and consider only the contribution to v from the hydrostatic pressure term, so that from (a 20)

$$v_0(0) = \frac{c}{r}\sin\theta \propto \frac{a}{r}\sin\theta, \tag{a 25}$$

where a is the 'strength' of the dislocation, i.e. the interatomic spacing provided with a ± sign according to the sign of the dislocation. If the dislocation is oscillating along its slip-plane (the x-axis) with amplitude d, then at points where $r \gg d$

$$v_0(t) \propto \frac{a}{r}\sin\theta + d\sin\omega t\frac{\partial}{\partial x}\frac{a}{r}\sin\theta = \frac{a}{r}\sin\theta - \frac{ad}{r^2}\sin 2\theta\sin\omega t. \tag{a 26}$$

The constant first term does not contribute to the damping. When the sign of a is changed the direction of motion under a given applied shear is reversed; in other words, the sign of d is also changed, so that the second term in (a 26) is unaltered. This is illustrated in figure 1 (*a*) and (*b*), where the T's represent the slip-plane and extra half-plane of atoms of an edge dislocation. The arrow denotes the direction of motion under the shear σ. The rule is that the dislocation moves in the direction of the σ-arrow to which the stem of its T points. The ± signs give the sign of the second term in (a 26) in the four quadrants. If the slip-plane of the dislocation is turned through a right angle, (*c*) or (*d*), the second term of (a 26) is unaltered except possibly in sign. The stresses σ are accompanied by the stresses σ' on account of the symmetry of the stress-tensor. The diagrams (*c*) and (*d*) can now be filled in as were (*a*) and (*b*). It is clear that at large distances from their centres the four dislocations give indistinguishable temperature variations. In this sense they are all of the same sign. Of course at distances of the order of d from their centres the temperature distribution about (a), (b), (c) and (d) are different. If the damping due to a pair of dislocations in a given relative position is calculated in detail the result is found to be independent of whether the pair is of type (aa), (ab), (ac), ..., provided the distance between them is large compared with d.

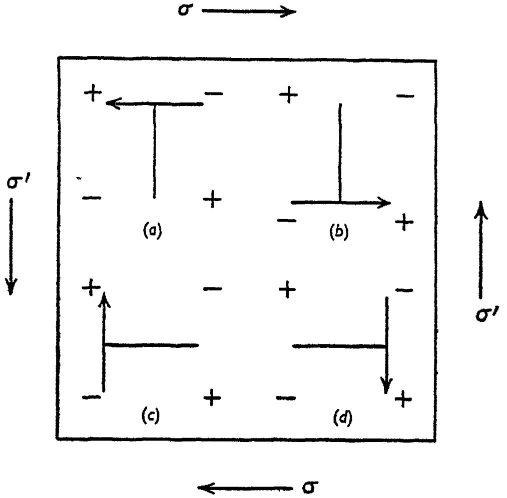


FIGURE 1.

We shall therefore assume that a good approximation to the actual case of a large number of edge dislocations will be obtained by attributing to them the $v_0(0)$ of (a 25). In addition to taking them all of the same sign we also assume that d is the same for all of them.

For a number N of identical dislocations in an infinite medium, situated at the points $\mathbf{r}_n$, we have by (a 11)

$$\overline{V}_0 = \bar{v}_0 \sum_0^N e^{i\mathbf{k}.\mathbf{r}_n}, \tag{a 27}$$

where $\overline{V}_0$ is the transform for the set of dislocations and $\bar{v}_0$ refers to a single one at the origin. Then

$$\begin{aligned} \overline{V}\,\overline{V}_0^* &= \overline{v}\overline{v}_0^* \sum_{m,n} e^{i\mathbf{k}.(\mathbf{r}_m-\mathbf{r}_n)} \\ &= \overline{v}\overline{v}_0^*\{N + 2 \sum_{m>n} \cos \mathbf{k}.(\mathbf{r}_m - \mathbf{r}_n)\}, \end{aligned} \tag{a 28}$$

where there are $\frac{1}{2}N(N-1)$ terms in the summation. Suppose that the N dislocations are scattered over a circle of radius r_1. We now assume that the summation in (a 27) may be replaced by the integral

$$\tfrac{1}{2}N(N-1) \int_{\mathbf{r}} \frac{d\mathbf{r}}{S} \int_{\mathbf{r}'} \frac{d\mathbf{r}'}{S} \cos \mathbf{k}.(\mathbf{r}-\mathbf{r}')\, P(\mathbf{r}-\mathbf{r}'), \tag{a 29}$$

where $S = \pi r_1^2$ is the area of the circle and $P(\mathbf{r}-\mathbf{r}')$ is a density function expressing the relative probability of finding a dislocation at a vector distance $\mathbf{r}-\mathbf{r}'$ from any particular dislocation. For a completely random distribution $P = 1$ everywhere inside r_1. If the dislocations tend to avoid one another and form an irregular lattice, P will be less than unity for small $|\mathbf{r}-\mathbf{r}'|$ and will approach unity for large $|\mathbf{r}-\mathbf{r}'|$. As a rough approximation to a more complicated distribution function we take

$$\begin{aligned} P &= 1, \quad |\mathbf{r}-\mathbf{r}'| \geqslant r_0, \\ P &= 0, \quad |\mathbf{r}-\mathbf{r}'| < r_0, \end{aligned}$$

i.e. each dislocation is surrounded by a circle of radius r_0 inside which another cannot lie. It is easiest to carry out the integration first assuming that $P = 1$ everywhere and then subtract a correction. The integral in (a 29) then becomes

$$\int_0^{2\pi}\int_0^{r_1}\int_0^{2\pi}\int_0^{r_1} \cos \mathbf{k}.(\mathbf{r}-\mathbf{r}')\, r\,dr\,d\theta\, r'\,dr'\,d\theta', \tag{a 30}$$

less the correction $$\int_0^{2\pi}\int_0^{r_1}\int_0^{2\pi}\int_0^{r_0} \cos \mathbf{k}.(\mathbf{r}-\mathbf{r}')\, r\,dr\,d\theta\, r'\,dr'\,d\theta', \tag{a 31}$$

with the understanding that in the integration from 0 to r_0 only those parts of the circle of radius r_0 about $\mathbf{r}$ which lie within r_1 are to contribute. If

$$I(x) = \int_0^{2\pi}\int_0^{x} \cos(kr\cos\theta)\, r\,dr\,d\theta = 2\pi x^2 \frac{J_1(kx)}{kx},$$

the expression (a 30) is $I^2(r_1)$. In (a 31) the integration over the dashed co-ordinates is equivalent to multiplication by πr_1^2 with a small error of order $r_0 r_1/r_1^2$ due to the intersection of some of the circles of radius r_0 with the circumference of r_1. Integration over the undashed co-ordinates yields $I(r_0)$. Hence if we replace $N(N-1)$ by N^2 (a 28) becomes

$$\overline{V\dot{V}_0^*} = \overline{v\dot{v}_0^*}\, N\left\{1 + 4N\frac{J_1^2(kr_1)}{(kr_1)^2} - 2N\frac{r_0^2}{r_1^2}\frac{J_1(kr_0)}{kr_0}\right\}.$$

The energy loss will be given by (a 23) with the factor $N\{\ \}$ inserted in the integrand. The second term in $\{\ \}$ represents, for large r_1, a peak at the origin of height N and width of the order of $1/r_1$. It can easily be shown to give a negligible contribution to w. Since $|J_1(x)/x| \leqslant \frac{1}{2}$ the w of (a 23) will be altered by a factor lying between the limits $N(1 \pm Nr_0^2/r_1^2)$. $\pi r_1^2/N$ is the mean area occupied per dislocation and πr_0^2 is the forbidden area surrounding each dislocation; Nr_0^2/r_1^2 will therefore be a small fraction of unity for small departures from complete randomness, and the loss per cycle will be approximately N times that of a single dislocation. In the extreme case where the circles r_0 round each dislocation form a two-dimensional close-packed array $Nr_0^2/r_1^2 = \frac{1}{2}\pi\sqrt{3} \doteqdot 0{\cdot}9$, and a detailed calculation is necessary. If $N\{\ \}$ is inserted in the integrand of (a 23) only the first term of the expansion of H_3 need be retained if $r_0 \gg l$. Using the result

$$\int_0^\infty \frac{J_1(x)}{x^4+a^4}\, x^2\,dx = (\mathrm{kei}'\,a)/a,$$

we easily find $$w = N\frac{\pi^2 c_p d'\omega C^2}{2T\kappa}\left\{\tfrac{1}{4}\log\frac{\kappa}{\omega l^2} - N\frac{r_0^2}{r_1^2}\frac{\mathrm{kei}'\,[r_0(\omega/\kappa)^{\frac{1}{2}}]}{r_0(\omega/\kappa)^{\frac{1}{2}}}\right\}.$$

For $x > 1$ $\mathrm{kei}'\,x$ falls off exponentially with x, whilst for small x, $x^{-1}\mathrm{kei}'\,x \sim -\frac{1}{2}\log x$. Thus when $r_0 > (\omega/\kappa)^{\frac{1}{2}}$, w is effectively N times that of a single dislocation, and when $r_0 < (\omega/\kappa)^{\frac{1}{2}}$ the same result holds with $\log(\kappa/\omega l^2)$ replaced by $\log(r_0^2/l^2)$. For kilocycle frequencies and $l \sim 10^{-7}$ cm., $\log(\kappa/\omega l^2) \sim 20$, and for any reasonable value of r_0, $\log(r_0^2/l^2)$ will be of the same order.

The above calculations refer to a finite collection of dislocations in an infinite medium. The physical case is that of a finite body filled with dislocations. Imagine the finite body inserted into a suitable cavity in an infinite medium. The oscillating heat-sources representing the dislocations near the surface of the original body will only affect the new material to a depth of the order of $(\kappa/\omega)^{\frac{1}{2}}$ (cf. Carslaw & Jaeger 1947). Since the process of energy-loss is essentially a volume effect the damping will only be altered by a fraction of the order of $L^2(\kappa/\omega)^{\frac{1}{2}}/L^3$, where L is some linear dimension of the body. Provided $(\kappa/\omega)^{\frac{1}{2}} \ll L$ the results for the infinite case may be applied to the finite body.

A 6. *Radiation damping of dislocations*

For simplicity consider a screw dislocation. The stress around it is everywhere a pure shear. If the centre of the dislocation is at $(\xi, 0)$, then by (a 13) the displacement is

$$u_3 = \frac{a}{2\pi}\arctan\frac{y}{x-\xi} \div \frac{a}{2\pi}\theta + \frac{a}{2\pi}\frac{\sin\theta}{r}\xi \tag{a 32}$$

for $r \gg \xi$. If we put $\xi = d\cos\omega t$ in (a 32), u_3 will not be a solution of the elastic wave-equation. It will be approximately correct for r much less than the wave-length of sound of frequency $\omega/2\pi$ in the material. The problem is to find a solution of the wave-equation representing outgoing waves for large r and approximating to (a 32) with $\xi = d\cos\omega t$ for small r. Since the term $a\theta/2\pi$ already satisfies the wave-equation the problem reduces to solving

$$c^2\nabla^2 u - \ddot{u} = 0, \quad c = (G/\rho)^{\frac{1}{2}},$$

with

$$u = \frac{ad}{2\pi}\frac{\sin\theta}{r}\mathrm{e}^{-i\omega t} \quad (d \ll r \ll 1/k)$$

and

$$u = F(r,\theta)\,\mathrm{e}^{i(kr-\omega t)} \quad (r \gg 1/k),$$

where c is the velocity of shear waves and $k = \omega/c$. The solution must be of the form

$$u = A\,\mathrm{e}^{-i\omega t}\sin\theta\, Z_1(kr),$$

where Z_1 is a Bessel function of order unity. The Bessel function which behaves like e^{ix} for large x is $H_1^{(1)}(x)$, which has for large and small x the forms $2i/\pi x$ and $(2/\pi x)^{\frac{1}{2}}\mathrm{e}^{ix}$ respectively. This easily leads to the following form for u at large r, expressed now in real form:

$$u = \frac{akd}{4}\left(\frac{2}{\pi kr}\right)^{\frac{1}{2}}\cos[\phi(r) - \omega t]\sin\theta.$$

The rate of flux of energy in the direction of propagation (radially outwards) is $\frac{1}{2}\rho u^2 c$ per unit area (Rayleigh 1894). If this is integrated over a cylinder of radius r and over a period of the oscillation it is easily shown that

$$w = \frac{\pi}{16}\rho a^2 d^2 \omega^2 \tag{a 33}$$

is the energy radiated away per cycle per unit length of dislocation.

REFERENCES

Burgers, J. M. 1939 *Proc. K. Akad. Wet. Amst.* **42**, 293.
Carslaw, H. S. & Jaeger, J. C. 1947 *Conduction of heat in solids.* Oxford: Clarendon Press.
Debye, P. 1927 *Phys. Z.* **28**, 135.
Kauzmann, W. 1941 *Trans. Amer. Inst. Min. (Metall.) Engrs*, **143**, 57.
Koehler, J. S. 1941 *Phys. Rev.* **60**, 397.
Lawson, A. W. 1941 *Phys. Rev.* **60**, 330.
Mott, N. F. & Nabarro, F. R. N. 1948 *Phys. Soc. Report of Conference on the Strength of Solids*, p. 1.
Päsler, M. 1944 *Z. Phys.* **122**, 357.
Rayleigh, Lord 1894 *Theory of sound.* London: Macmillan.
Read, T. A. 1941 *Trans. Amer. Inst. Min. (Metall.) Engrs*, **143**, 30.
Read, T. A. & Tyndall, P. T. 1946 *J. Appl. Phys.* **17** (9), 720.
Seitz, F. 1943 *The physics of metals.* New York: McGraw-Hill.
Seitz, F. & Read, T. A. 1941 *J. Appl. Phys.* **12** (6), 475.
Swift, L. M. & Richardson, J. E. 1947 *J. Appl. Phys.* **18** (4), 417.
Titchmarsh, E. C. 1937 *Fourier integrals.* Oxford: Clarendon Press.
Watson, G. N. 1922 *A treatise on Bessel functions.* Cambridge University Press.
Zener, C. 1940 *Proc. Phys. Soc.* **52**, 152.

The determination of static and dynamic yield stresses using a steel ball

BY R. M. DAVIES, *University College of Wales, Aberystwyth*

(*Communicated by Sir Geoffrey Taylor, F.R.S.—Received* 7 *December* 1948)

The static and dynamical yield stress of the material of a thick steel plate may be estimated by pressing and by dropping a hard steel ball on a plane surface of the plate which has been ground and then polished. Under these conditions, the first appearance of an indentation on the polished surface can be detected with good accuracy, either by an optical interference method, or by an optical reflexion method. The statical experiment consists in finding the least force which must be applied to the steel ball to produce a permanent indentation, whilst the dynamical experiment consists in finding the least normal velocity of impact which gives similarly a permanent indentation. Using either the Guest-Mohr principal-stress difference or the von Mises shear strain energy hypotheses as criteria of failure, combined with an analysis of the stresses in the plate, it is shown how the appropriate yield stress can be calculated from the experimental data.

Tests were made on a specimen of mild steel, two specimens of homogeneous armour plate and a very hard nickel-chrome steel of the type used for ball and roller bearings. The ratio of the dynamic value of the yield stress to the static value was found to increase as the hardness number decreases; the ratio was practically unity for the nickel-chrome steel, about 1·1 for the armour plate and about 2 for the mild steel.

The values of the static yield stress found by the ball method and by an ordinary tensile or compression test are different; this is probably due partly to the inaccuracy of the criteria of plastic flow, partly to the difference in work-hardening in the two experiments, and partly to changes in the structure of the surface due to polishing. This discrepancy is without effect on the ratio of the dynamic to static yield stress as determined by the ball method, since the stress distributions in the static and dynamic ball experiments are identical.

1. INTRODUCTION

In the early part of 1941, a series of experiments was carried out at the Engineering Laboratory, Cambridge, in conjunction with Sir Geoffrey Taylor, to estimate the static and dynamic yield stresses of the material of a thick steel plate by pressing and dropping a steel ball on the plate. A brief account of the work has already been given by Sir Geoffrey in the 'James Forrest Lecture' to the Institution of Civil Engineers (Taylor 1946). When a steel ball is pressed with a gradually increasing force against the surface of a steel plate, the behaviour of the plate and ball is elastic up to a certain limit determined by the stress distribution in the system; beyond this limit, plastic flow occurs and if, as is usually the case, the hardness of the ball exceeds that of the plate, failure will be accompanied by the appearance of a permanent indentation in the surface of the plate. If this surface is first ground plane and then polished, the first appearance of an indentation can be detected with considerable accuracy by optical methods, and thus the minimum load, $P_{\text{min.}}$ say, required to produce plastic flow can be determined experimentally.

Similarly, in the impact of a steel ball against a steel plate, the behaviour is elastic for small velocities, but at greater velocities, the plastic limit is reached and, as in the static case, it is possible to determine experimentally the minimum velocity of impact, $v_{\text{min.}}$ say, which gives rise to a permanent deformation.

Knowing $P_{\text{min.}}$ and $v_{\text{min.}}$, the corresponding stress distributions at the plastic limit in the material of the plate can be calculated, and by applying one of the recognized empirical criteria of failure, the yield stress can be calculated. In the case of ductile materials, such as steel, for which the tensile yield stress is approximately equal to the compressive yield stress, the Guest-Mohr hypothesis appears to be reasonably satisfactory as a criterion of failure, and it is adopted as such in this paper. According to this hypothesis, failure first sets in where and when the maximum value of the difference of the so-called principal stresses becomes equal to the value of the direct yield stress determined in the usual way in a tensile test. It thus becomes necessary to calculate the differences between the various principal stresses in the case of a sphere pressing or impinging against the plane surface of an elastic semi-infinite medium.

The fundamental analysis of these problems in the elastic regime was first given by Hertz (1881); the relevant results for the purpose of the present paper are given in § 2 below. Hertz's investigations were confined to the phenomena occurring in the area of contact of the bodies, although his equations apply to the whole of the bodies in contact; explicit solutions for the internal stresses and strains, derived from the fundamental equations given by Hertz (1881), have been given by Huber (1904), by Fuchs (1913), and by Morton & Close (1922) for the case where the area of contact is a circle, whilst Thomas & Hoersch (1930) have given a more general solution covering the case where the area of contact is not a circle. Numerical calculations, based on Huber's equations and applicable to the present work, are given in § 3 below. For the contact of a sphere and a plate, the most striking result is that, whereas the actual value of the normal pressure is a maximum at the centre of the area of contact, the difference of the principal stresses reaches a maximum slightly below the centre of the area of contact, in the material of the plate. After this result had been found, it was discovered that it had been noted, for the somewhat similar case of a thick plate under a load distributed uniformly over a circle, by Coker & Filon (1931) and by Thomas & Hoersch (1930) for the more general case.

From this result it follows, according to the Guest-Mohr hypothesis, that failure will start, not in the area of contact, but at a point below this. In point of fact this has been proved to be the case in a series of experiments by Thomas & Hoersch (1930); these experiments were carried out with large case-hardened steel cylinders pressed into contact with mild steel plates with forces of the order of 10 tons. The flow which occurred when the plastic limit was exceeded was detected by sectioning the steel plates and etching the sections by a process due to Fry (1921).

According to the Guest-Mohr hypothesis, the maximum difference of principal stresses at the plastic limit is equal to the yield stress determined in the usual way by a tensile test; for steel with Poisson's ratio equal to 0·286, the calculations given below show that the maximum value of the difference of the principal stresses is equal to 0·63 times the maximum normal pressure in the area of contact, and since the value of this quantity at the plastic limit can be determined from the values of $P_{\text{min.}}$ and $v_{\text{min.}}$, it follows that the yield stresses under static and dynamic conditions can be calculated. In connexion with the dynamical experiments, it should be pointed out that the rate of stressing is extremely high; the time of contact of ball and the plate is of order 30μsec. (1μsec. $= 10^{-6}$ sec.); the time of application of stress is one-half of this, whilst the maximum difference of principal stresses at the yield point is of order 81 tons/in.^{2}; the average rate of stressing is thus of order $5{\cdot}4\times10^6$ tons/in.^{2} sec.

It should be pointed out that there is some similarity in experimental procedure between the static experiments described in this report and some experiments (notably by Hertz and Auerbach) on hardness; an excellent summary of these experiments is given by O'Neill (1934). A fundamental difference does however exist, since previous experiments (apart from those of Thomas & Hoersch 1930), have confined themselves to a consideration of the stresses in the area of contact. No trace of any account of an attempt to measure the dynamical yield stress by a ball method has been found in the literature.

Recently, an interesting paper on the related subject of the theory of static and dynamic hardness has been published by Tabor (1948); like most investigators of hardness, Tabor is concerned with the fully developed plastic flow which occurs in practical hardness tests, whereas the present paper deals with the first observable transition from the elastic to the plastic states.

2. STRESSES IN THE AREA OF CONTACT OF A SPHERE AND A SEMI-INFINITE SOLID

The theory of the stresses and strains produced when an elastic sphere is pressed statically against the plane surface of a semi-infinite elastic body is a special case of Hertz's theory of contact between two elastic spheres. Considering the case of a sphere of radius R pressed with a total force P against the plane surface of a semi-infinite body of the same material, then under equilibrium conditions, the sphere and the solid are in contact over a portion of a spherical surface of radius $2R$; the surface of contact is bounded by a circle—the circle of contact—and if the radius of this circle is a, then by a reduction of Hertz's equations as given by Love (1934), it can be shown that

$$a^3 = 1{\cdot}5(1-\sigma^2)\,PR/E, \tag{2·1}$$

where E and σ are respectively the Young's modulus and the Poisson's ratio of the material of the sphere and solid.

The distribution of normal pressure over the area of contact follows the law

$$p = p'\sqrt{(a^2-r^2)}/a, \tag{2·2}$$

where p is the normal pressure at distance r from the centre of the circle of contact and p' is the maximum normal pressure in the circle of contact. Clearly $p = p'$ when $r = 0$. From equation (2·1) the mean normal pressure, $\overline{p}$, over the area of contact is

$$\overline{p} = \frac{p}{\pi a^2} = \frac{0{\cdot}7634}{\pi}\left(\frac{E}{1-\sigma^2}\right)^{\frac{2}{3}}\left(\frac{P}{R^2}\right)^{\frac{1}{3}}. \tag{2·3}$$

From equation (2·2), it can be shown that

$$p' = 1{\cdot}5\overline{p}, \tag{2·4}$$

and thus

$$p' = \frac{1{\cdot}145}{\pi}\left(\frac{E}{1-\sigma^2}\right)^{\frac{2}{3}}\left(\frac{P}{R^2}\right)^{\frac{1}{3}}. \tag{2·5}$$

For steel, we may assume $E = 20\times10^{11}$ dynes/cm.^{2} = 12,950 tons/in.^{2}, $\sigma = 0{\cdot}286$ (cf. Southwell 1936); equations (2·1), (2·3) and (2·5) then become

$$a^3 = 6{\cdot}88_5\times10^{-13}PR, \tag{2·1a}$$

$$\overline{p} = 4{\cdot}09_7\times10^{7}\,\frac{P^{\frac{1}{3}}}{R^{\frac{2}{3}}}, \tag{2·3a}$$

$$p' = 6{\cdot}14_0\times10^{7}\,\frac{P^{\frac{1}{3}}}{R^{\frac{2}{3}}}, \tag{2·5a}$$

where P is expressed in dynes, $\overline{p}$ and p' in dynes/cm.^{2} and R in cm.

In order to gain some idea of the order of magnitude of the quantities involved, table 1 gives the total thrust P required to produce a maximum normal pressure of $1{\cdot}85\times10^{10}$ dynes/cm.^{2}, which equals 120 tons/in.^{2} (1 ton/in.^{2} = $1{\cdot}545\times10^{8}$ dynes/cm.^{2}) for steel balls of various diameters; the last column gives the radius of the circle of contact under these conditions.

TABLE 1

p' = 120 tons/in.^{2}

diameter of ball (in.)	P kg.	P lb.	a cm. ($\times10^{-3}$)	a in. ($\times10^{-3}$)
$\frac{1}{8}$	0·703	1·55	4·23	1·67
$\frac{1}{4}$	2·81	6·20	8·46	3·33
$\frac{3}{8}$	6·33	14·0	12·7	5·00

Turning now to the case of impact of elastic bodies at moderate velocities, the problems of elastic contact and elastic impact are in essence identical (Love 1934) and, allowing for the fact that the force between the bodies in impact is due to the deceleration of the moving bodies, the mechanics of the impact of a sphere of radius R, striking normally the plane surface of a semi-infinite solid with a velocity v, can be developed. If ρ is the density of the material of the sphere and the solid, then the duration of impact, τ, is given by

$$\tau = 2{\cdot}943\left[2{\cdot}5\pi\rho\left(\frac{1-\sigma^2}{E}\right)\right]^{\frac{2}{5}}\frac{R}{v^{\frac{1}{5}}}. \tag{2·6}$$

The maximum value, P_m, of the total force developed during impact is

$$P_m = \tfrac{2}{3}\,(2{\cdot}5\pi\rho)^{\frac{3}{5}}\left(\frac{E}{1-\sigma^2}\right)^{\frac{2}{5}} R^2 v^{\frac{6}{5}}. \tag{2·7}$$

At maximum compression, the radius, a_m, of the circle of contact is

$$a_m = \left[2{\cdot}5\pi\rho\left(\frac{1-\sigma^2}{E}\right)\right]^{\frac{1}{5}} R v^{\frac{2}{5}}. \tag{2·8}$$

Again, at maximum compression, the mean normal pressure, $\overline{p}_m$, is

$$\overline{p}_m = \frac{P_m}{\pi a_m^2} = \frac{2}{3\pi}(2{\cdot}5\pi\rho)^{\frac{1}{5}}\left[\frac{E}{1-\sigma^2}\right]^{\frac{4}{5}} v^{\frac{2}{5}}, \tag{2·9}$$

and the pressure at the centre of the circle of contact, p'_m (again at maximum compression), is

$$p'_m = 1{\cdot}5\overline{p}_m = \frac{1}{\pi}(2{\cdot}5\pi\rho)^{\frac{1}{5}}\left\{\frac{E}{1-\sigma^2}\right\}^{\frac{4}{5}} v^{\frac{2}{5}}. \tag{2·10}$$

For steel, with the values assumed above for E and σ, and with $\rho = 7{\cdot}8$ g./cm.^{3}, these equations become

$$\tau = 1{\cdot}77_5 \times 10^{-4} R/v^{\frac{1}{5}}, \tag{2·6a}$$

$$P_m = 6{\cdot}79 \times 10^5 R^2 v^{\frac{6}{5}}, \tag{2·7a}$$

$$a_m = 7{\cdot}76 \times 10^{-3} R v^{\frac{2}{5}}, \tag{2·8a}$$

$$\overline{p}_m = 3{\cdot}59 \times 10^9 v^{\frac{2}{5}}, \tag{2·9a}$$

$$p'_m = 5{\cdot}39 \times 10^9 v^{\frac{2}{5}}, \tag{2·10a}$$

where τ is in sec., a_m and R in cm., v in cm./sec., P_m in dynes, p'_m and $\overline{p}_m$ in dynes/cm.^{2}.

For comparison with experimental results, it is convenient to express these quantities in terms of the vertical height h through which the ball must be dropped in order to acquire velocity v. If h is in cm., then

$$\tau = 8{\cdot}31 \times 10^{-5} R/h^{\frac{1}{10}}\ \text{sec.}, \tag{2·6b}$$

$$P_m = 6{\cdot}42 \times 10^7 R^2 h^{\frac{3}{5}}\ \text{dynes}, \tag{2·7b}$$

$$a_m = 3{\cdot}54 \times 10^{-2} R h^{\frac{1}{5}}\ \text{cm.}, \tag{2·8b}$$

$$\begin{aligned}\overline{p}_m &= 1{\cdot}64 \times 10^{10} h^{\frac{1}{5}}\ \text{dynes/cm.}^2\\ &= 106 h^{\frac{1}{5}}\ \text{tons/in.}^2,\end{aligned} \tag{2·9b}$$

$$\begin{aligned}p'_m &= 2{\cdot}46 \times 10^{10} h^{\frac{1}{5}}\ \text{dynes/cm.}^2\\ &= 159 h^{\frac{1}{5}}\ \text{tons/in.}^2.\end{aligned} \tag{2·10b}$$

The values of these quantities for the elastic impact of a $\frac{1}{4}$ in. ball, for various values of v and h, are given in table 2.

TABLE 2

Diameter of ball $= 2R = \frac{1}{4}$ in. $= 0{\cdot}636$ cm.

h	v	τ	a_m		p_m	p'_m
(cm.)	(cm./sec.)	(μsec.)	(cm.)	(in.)	(tons/in.^{2})	(tons/in.^{2})
0·1	14·0	33·5	0·0071	0·0028	66·6	100
0·25	22·2	30·6	0·0085	0·0033	80·6	121
0·5	31·4	28·3	0·0098	0·0039	92·6	139
1	44·3	26·4	0·0113	0·0045	106	159
5	99·1	22·5	0·0155	0·0061	146	219
10	140	21·0	0·0178	0·0070	168	252
25	222	19·1	0·0214	0·0084	202	302
100	443	16·7	0·0282	0·0111	267	400

A striking feature brought out by the data given in this table is the surprisingly large values of the normal pressure given by small values of velocity and height of drop.

3. INTERNAL STRESSES DUE TO CONTACT OF A SPHERE AND A SEMI-INFINITE SOLID

For the calculation of the yield stress in the way outlined in § 1, a knowledge of the principal stresses in the interior of the semi-infinite solid is required. Figure 1 shows the sphere in contact with the semi-infinite solid; the radius of the circle of contact is a; the load, P dynes, is assumed to be vertical and the plane surface AB of the solid horizontal. For the sake of clearness, the deformations have been exaggerated. Let the point of contact of the sphere with the plane, when the load P is infinitesimally small, be taken as the origin, O, of a system of rectangular co-ordinates, the z axis being normal to the surface AB and the x and y axes lying in the surface. Because of symmetry about Oz, it is sufficient to consider stresses in the yz plane, which will be taken as the plane of the figure. Let C be the point (y, z), or (r, θ) referred to plane polar co-ordinates with O as origin and Oz as initial line. Because of

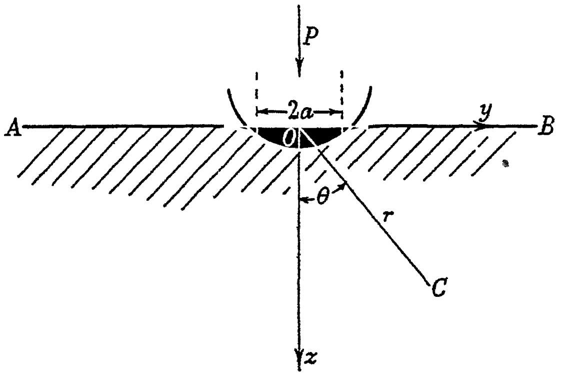


FIGURE 1

symmetry, the stresses at C reduce to normal stresses X_x, Y_y, Z_z and the tangential stress Y_z, where, in the usual notation, X_x is the force per unit area, in the direction of x, on an infinitesimal area with its normal along O_x, ...; Y_z is the force per unit area, in the direction of y, on an infinitesimal area with its normal along O_z.

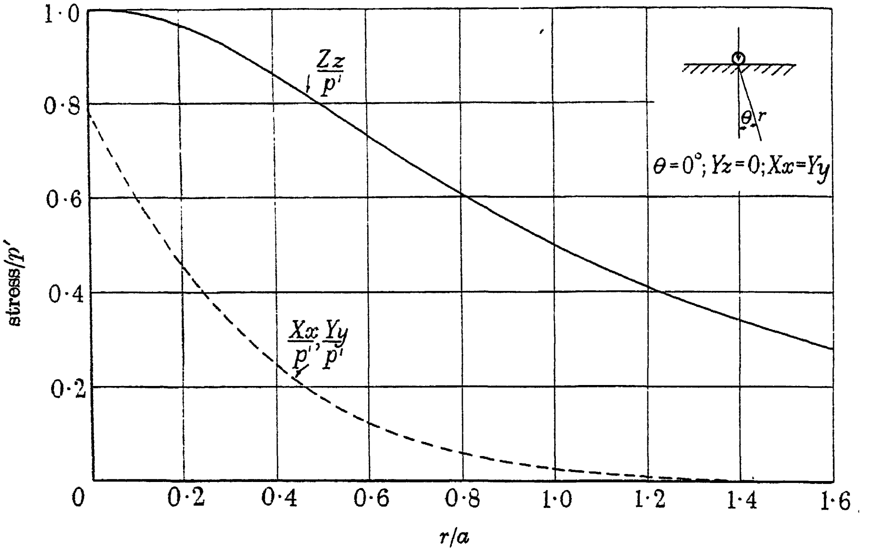


FIGURE 2

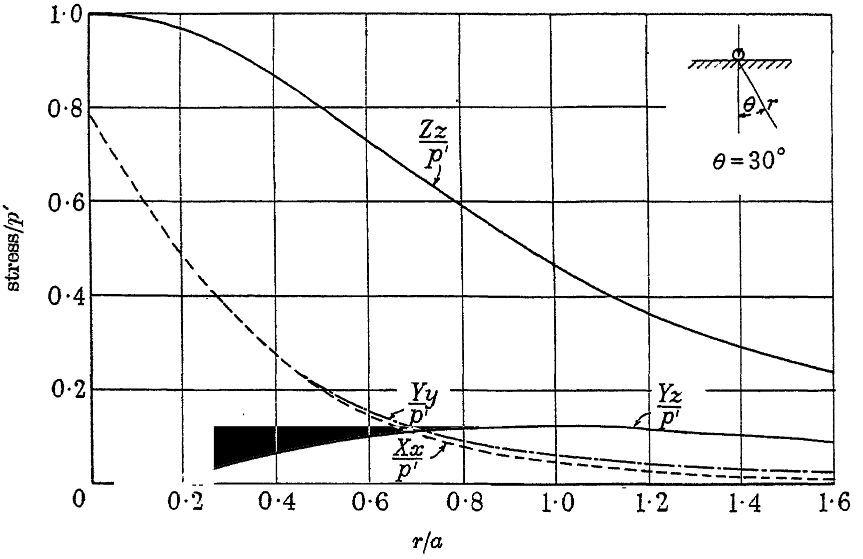


FIGURE 3

Numerical values for X_x, ..., Y_z for the contact of a sphere and a plane have been worked out for steel, with the elastic constants given above, from the equations of Huber (1904) and these are shown in figures 2 to 5 for $\theta = 0$, 30, 60 and 90°

respectively. In these figures, the ratio r/a (where a is the radius of the circle of contact) is taken as abscissa, and the ratios $X_x/p', \ldots, Y_z/p'$ (where p' is the maximum normal pressure in the circle of contact) are taken as ordinates.

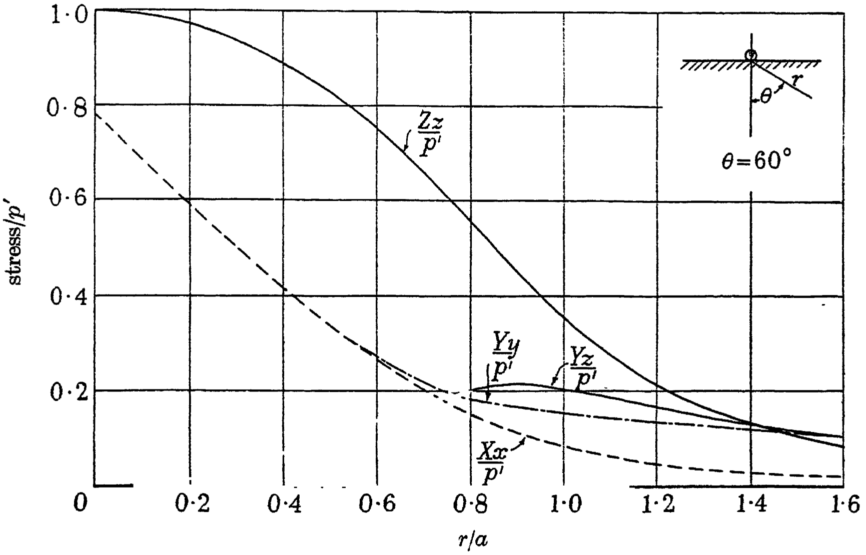


FIGURE 4

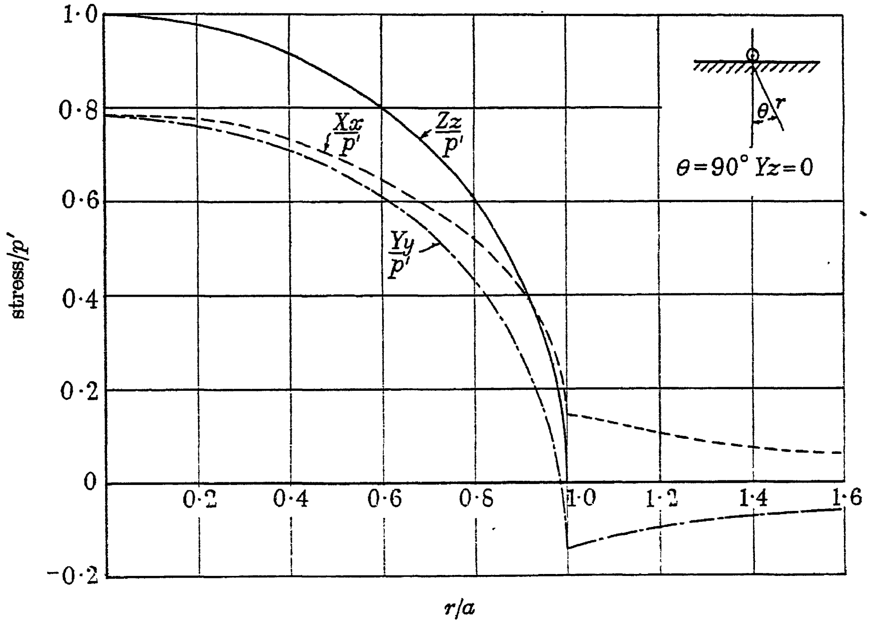


FIGURE 5

From the values of $X_x, \ldots, Y_z$, the values of the three principal stresses can be determined. Because of symmetry, the three principal planes of stress, i.e. the planes on which the shear stresses vanish, are the yz plane and two planes perpendicular to the yz plane and to each other; if γ, $\gamma+\frac{1}{2}\pi$ are the angles between these planes and Oy, then by well-known relations (Southwell 1936, p. 136, equations (27) and (29)), $\cot 2\gamma = (Z_z - Y_y)/2Y_z$, and if σ_1, σ_2 and σ_3 are the principal stresses, i.e. the normal stresses acting on the principal planes of stress, then

$$\sigma_1 = X_x; \quad \sigma_2, \sigma_3 = \tfrac{1}{2}(Y_y + Z_z) \pm \tfrac{1}{2}\sqrt{\{(Z_z - Y_y)^2 + 4Y_z^2\}}. \tag{3·1}$$

In the present case, except for $r \geqslant a$ when $\theta = 90°$, $|\sigma_2 - \sigma_3|$ is greater than either $|\sigma_3 - \sigma_1|$ or $|\sigma_1 - \sigma_2|$; apart from the region of the surface where $r \geqslant a$, it is $|\sigma_2 - \sigma_3|$ that determines the incidence of plastic flow in the material, and the variation of this quantity with r is shown in figure 6 for different values of θ; in this figure r/a is plotted as abscissa and $|\sigma_2 - \sigma_3|/p'$ as ordinate except for $\theta = 90°, r \geqslant a$, where the larger principal stress difference is taken instead of $|\sigma_2 - \sigma_3|$. The family of curves in figure 6 bring out clearly the point made in § 1, that the maximum value of the difference of the principal stresses occurs not in the area of contact but below this point, within

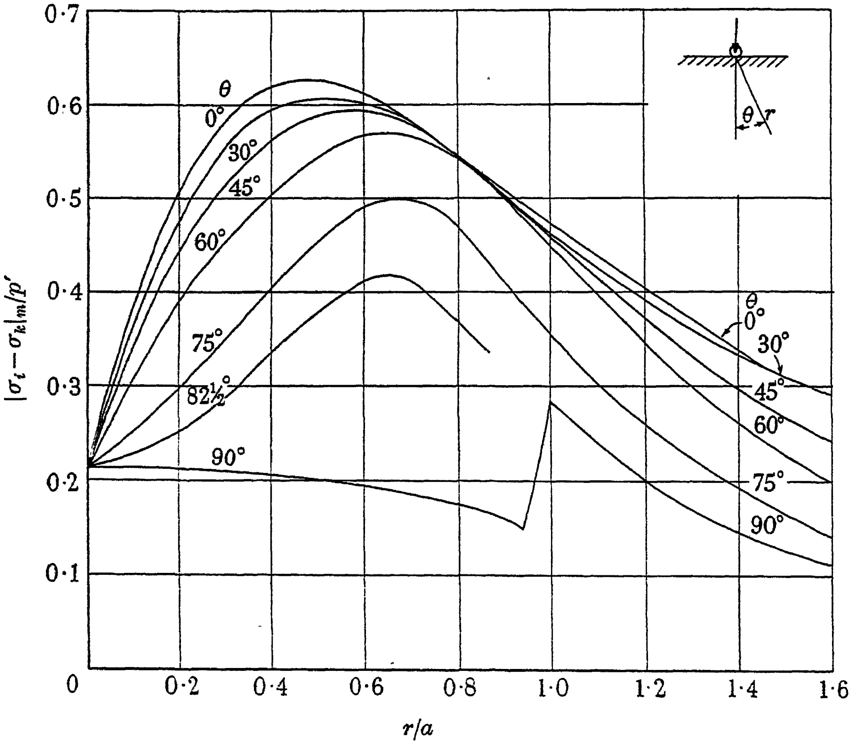


FIGURE 6. $|\sigma_i - \sigma_k|_m$ = largest difference of principal stresses.

the semi-infinite medium. The lines of constant principal stress difference in the interior of the semi-infinite solid are shown in figure 7, in which the surface is denoted by AA and the normal stress distributed over the surface of contact is represented by vertical lines drawn from AA to the semicircle CDC.

The maximum value of the difference of the principal stresses is $0{\cdot}63p'$ and therefore, according to the Guest-Mohr hypothesis, the yield stress, p_y say, of the material is equal to $0{\cdot}63p'_{\text{min.}}$, where $p'_{\text{min.}}$ is the value of the maximum normal pressure on the area of contact when permanent indentation first makes its appearance.

For a static experiment, from equation $(2{\cdot}5a)$, $p'_{\text{min.}} = 6{\cdot}14 \times 10^7 P^{\frac{1}{3}}_{\text{min.}}/R^{\frac{2}{3}}$, where, as in § 1, $P_{\text{min.}}$ is the least static load (expressed in dynes) required to produce an indentation. Thus, if p_{ys} is the static yield stress,

$$p_{ys} = 3{\cdot}87 \times 10^7 P^{\frac{1}{3}}_{\text{min.}}/R^{\frac{2}{3}} \text{ dynes/cm.}^2. \tag{3·2}$$

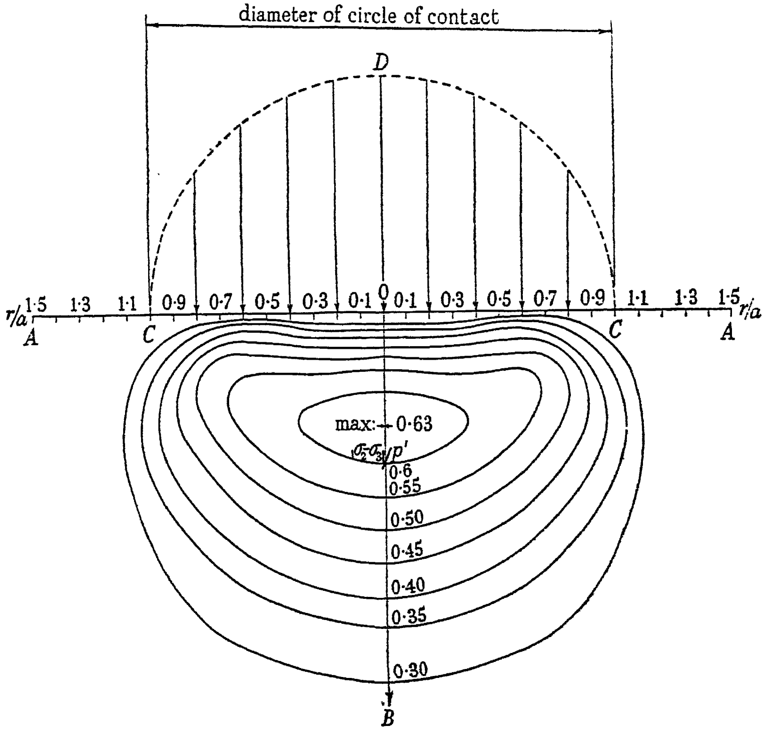


FIGURE 7

For a dynamic experiment, from equation $(2{\cdot}10a)$, $p'_{\text{min.}} = 5{\cdot}39 \times 10^9 v^{\frac{2}{5}}_{\text{min.}}$, where, as in § 1, $v_{\text{min.}}$ is the smallest normal velocity of impact (in cm./sec.) which gives rise to a permanent indentation. If p_{yd} is the dynamic yield point, then

$$p_{yd} = 0{\cdot}63p'_{\text{min.}} = 3{\cdot}40 \times 10^9 v^{\frac{2}{5}}_{\text{min.}} \text{ dynes/cm.}^2. \tag{3·3}$$

4. DESCRIPTION OF THE EXPERIMENTS

Tests were carried out on specimens of mild steel of low carbon content, two types of homogeneous armour plate (denoted by WTM and WTN respectively) and very hard nickel-chrome steel of the type used in roller bearings. The specimens varied in size and shape, but they were all more than $1\frac{1}{2}$ in. thick and their transverse dimensions were greater than 2 in. One face of each specimen was first ground plane and then polished so that over a small region (say, 1 cm. square) the surfaces were optically flat. When an indentation was made by a ball, statically or dynamically, on the polished surface of a test plate, the surface of the indentation appears to be spherical, and if the ball used for making the indentation was polished, the indentation was found to be sufficiently polished to reflect light like an optically worked mirror. Two optical methods were found to be available to detect an indentation and, if necessary, to measure its radius of curvature.

(a) *Interference method*

When a plane-parallel piece of optically worked glass was laid on the polished surface of the test plate, equidistant, straight, parallel interference fringes were seen by reflexion when the system was illuminated, using a sloped piece of glass, with monochromatic light. The fringes were observed in the usual way through a low-power microscope fitted with an eyepiece containing a cross-wire which was moved by a micrometer screw. The source of monochromatic light was either a sodium flame or a mercury arc fitted with a filter to isolate the green line; the wave-length of the light was $5{\cdot}893 \times 10^{-5}$ cm. in the former case and $5{\cdot}461 \times 10^{-5}$ cm. in the latter.

When an indentation was made in the surface of the test plate and the interference fringes were observed, the indented portion of the surface gave rise to one or more circular fringes in precisely the same way as the familiar Newton's rings formed by the curved surface of a spherical lens and an optical flat. These rings were easily detected in the field of view of the microscope, even if only one was present; if two or more were present, the radius of curvature of the indentation could be measured since two adjacent rings represent contour lines of the indentation differing in depth by $\frac{1}{2}\lambda$, where λ is the wave-length of the monochromatic light used. If x' and x'' are the radii of two consecutive dark fringes, and if ρ be now the radius of curvature of the surface, then

$$\rho = (x'^2 - x''^2)/\lambda. \tag{4·1}$$

If only one fringe was present, an estimate of the radius of curvature could be made by gently moving the upper glass plate, keeping the system of fringes parallel to their original direction, so that the circular ring was first in the centre of a bright fringe and secondly in the centre of the next dark fringe. The two radii of the ring (y' and y'', say) correspond to a difference in depth of $\frac{1}{4}\lambda$ and the radius of curvature, ρ, of the surface is

$$\rho = 2(y'^2 - y''^2)/\lambda. \tag{4·2}$$

(b) *Mirror method*

In this method, the indented portion of the surface was regarded as a minute spherical mirror. If the light from an illuminated small aperture at an infinite distance away were allowed to fall on this mirror, an image of the aperture would be formed at distance f from its pole, where f is the focal length of the mirror; if ρ is its radius of curvature, then $\rho = 2f$. In practice, the illuminated small aperture at infinity was a circular hole about 0·2 mm. diameter, placed at the focus of a converging lens and illuminated by the light from a 500 candle-power 'Pointolite' lamp.

The light from the lens travelled horizontally, and was reflected vertically on the horizontal polished surface of the steel plate by a microscope slide at 45° to the horizontal. The illuminated surface of the plate was viewed through a microscope with its axis vertical. When the microscope was focused on the surface, any indentation in the field of view became visible, and, as a rule, inevitable particles of fine dust were also seen. If the microscope was focused a little below the surface, the indentation appeared as a dark circular patch, surrounded by a bright circular ring, caused by the transition between the indented and plane portions of the polished surface. On racking the microscope up from the position in which it was focused on the surface, dust particles gave rise to various diffraction patterns, including a bright point of light surrounded by alternate bright and dark diffraction rings; an indentation similarly gave rise to various systems of circular diffraction rings, but at one fairly well-defined position of the microscope, the image of the circular aperture was clearly seen in focus as a very brilliant small circle of light. It was easy to distinguish between the image of a genuine indentation and the spurious images due to dust particles, since, on focusing on or slightly below the surface, the dust particles became clearly visible whilst the indentation gave the characteristic dark circular patch, surrounded by a bright circular ring.

Where necessary, the radius of curvature of the indentation was measured by finding the vertical distance through which the microscope must be moved from the position in which the polished surface is in focus to the position in which the image of the aperture is in focus.

Both methods of detecting an indentation are very sensitive and they will detect an indentation the depth of which is of the order of one wave-length of light.

Turning to the procedure adopted in applying a given load in statical experiments and in producing an impact with a given velocity in dynamical experiments, it is clearly necessary to arrange that contact between the ball and the plane surface takes place at an undamaged portion of the surface and that some method is available to ensure that the region of contact in a particular test can be identified for examination after the test. In the present experiments, a simple procedure for locating the test region relative to suitable rectangular co-ordinate axes, was used. It is also important that the ball and the plate should be degreased, by washing with a solvent such as carbon tetrachloride, before each test.

The apparatus used to apply the load in statical experiments is shown in figure 8, where A represents the steel block under test, resting on a sheet of rubber B and supported by the projecting portion of a length of 3 in. steel channel which was firmly clamped to a table. The steel ball D rested in a conical seating in the upper horizontal member E of the metal framework F; the load was applied by gradually and carefully adding weights to a scale pan attached to the mid-point G of the lower horizontal member of F, a spring S being interposed between G and the scale pan.

It is important that the ball should not roll on the surface of A when the load is being applied; to prevent rolling, E is kept horizontal during the application of the load, a preliminary balancing experiment having been made to ensure that the point G lies vertically under the centre of the conical seating at D when E is horizontal.

In the dynamical tests, an obvious way of carrying out the experiment would be to drop the ball vertically from the appropriate height on the horizontal surface of the plate. This method was unsuitable for the present experiments, since the heights of fall are quite small on account of the large stresses produced by small impact velocities, and it would therefore be difficult to catch the ball so as to avoid more than one impact. It was therefore decided to arrange the block with its polished face in a vertical plane, to suspend the ball on a fine thread, in pendulum fashion, and to allow it to fall, from a displaced position, so as to strike the block normally. When the length of the pendulum is 30 cm. a vertical drop of 1 cm., giving a velocity of impact of 44·3 cm./sec., corresponds to a horizontal displacement of about 7·5 cm., and it was easy to catch the ball by hand after the first impact.

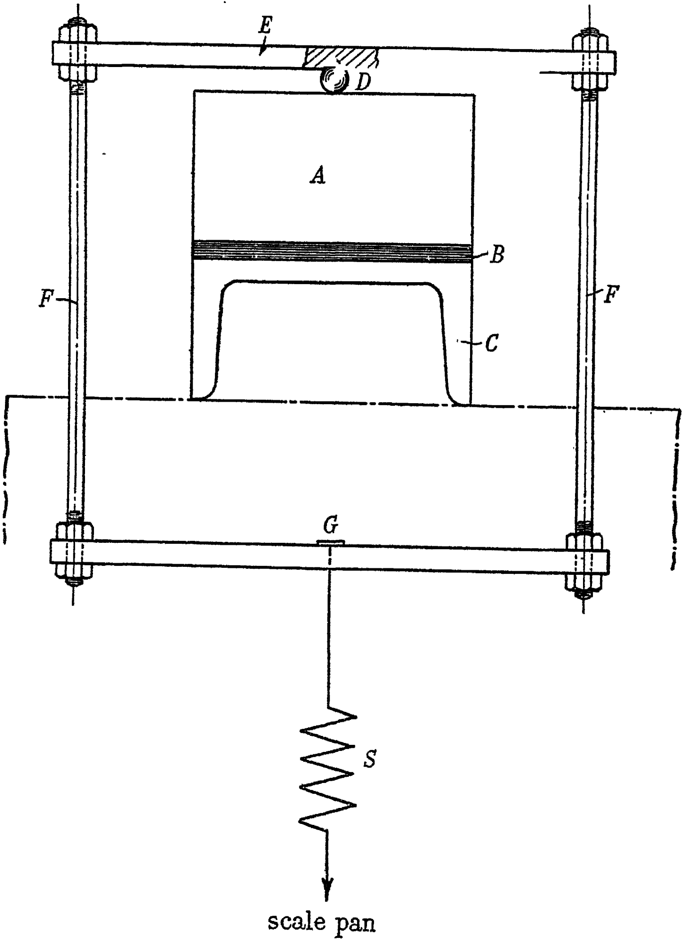


FIGURE 8

5. EXPERIMENTAL RESULTS

(a) *Hardness and tensile test results*

The results of tests on the specimens are summarized in table 3.

The figures given in the last two columns were obtained by ordinary tensile tests for the two specimens of armour plate and by a compression test for mild steel. The ultimate stress is calculated on the reduced area at fracture.

Table 3

specimen	Brinell number	Vickers pyramid number	yield stress (tons/in.^{2})	ultimate stress (tons/in.^{2})
mild steel	125	—	18	—
armour plate WTM	351	341	74	125
armour plate WTN	321	314	69	153
nickel-chrome steel	—	771	—	—

(*b*) *Static tests by the ball method*

The consistency of the results obtained by the ball method under static loading can be gauged from table 4, which relates to the specimens WTN of armour plate. In this table, the 'upper limiting load' is used to denote the smallest load which formed a permanent indentation, whilst the 'lower limiting load' denotes the largest load which produced no indentation; the mean of these two quantities is taken as the minimum load, $P_{\text{min.}}$, required to produce an indentation. As in § 3, $p'_{\text{min.}}$ is the maximum normal pressure in the area of contact when the total load is $P_{\text{min.}}$.

Table 4

	specimen WTN		
ball diameter (in.)	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{8}$
upper limiting load (kg.)	2·40	1·11	8·02
lower limiting load (kg.)	1·91	0·89	6·33
$P_{\text{min.}}$ (kg.)	2·15	1·00	7·18
$p'_{\text{min.}}$ (tons/in.^{2})	110	135	125
yield stress ($=0·63p'_{\text{min.}}$) (tons/in.^{2})	70	85	79

Table 4 shows that there is no serious systematic variation in the result for the yield stress when ball diameter is varied. The scatter of the results is due partly to experimental error and partly to non-uniformity in the plates, and to the effect of polishing on the structure of the surface layers. In this connexion, it must be remembered that the volume of metal affected by a test is very small, being of the order of the volume of a hemisphere whose radius is equal to the radius of the circle of contact given by Hertz's theory, i.e. of the order of 0·002 to 0·005 in.

(*c*) *Dynamic tests by the ball method*

The results of some experiments, carried out to test the consistency of the method, are given in table 5. As in § 2, h represents the vertical height through which the ball falls before impact; the 'upper limit of h' is the smallest value of h which was

Table 5

	specimen WTN		
ball diameter (in.)	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{8}$
upper limit of h (cm.)	0·4	0·5	0·4
lower limit of h (cm.)	0·3	0·4	0·3
$h_{\text{min.}}$ (cm.)	0·35	0·45	0·35
$p'_{\text{min.}}$(tons/in.^{2})	129	135·5	129
yield stress ($=0·63p'_{\text{min.}}$) (tons/in.^{2})	81·2	85·5	81·2
corresponding static yield stress (tons/in.^{2})	70	85	79

found to make an indentation, whilst the 'lower limit of h' is the largest value of h which was found to produce no indentation. The mean of these two figures was taken as h_{min}, the minimum value of h which gives an indentation. $p'_{\text{min.}}$ is the maximum normal pressure, in the area of contact, corresponding to h_{min}.

The results obtained with the three sizes of balls show that $v_{\text{min,}}$ is sensibly independent of ball diameter as predicted by theory (see equation (3·3)).

(*d*) *Summary of experimental results*

The results of the experiments carried out on the various specimens are summarized in table 6.

TABLE 6

		armour plate		
	mild steel	WTM	WTN	nickel-chrome steel
Brinell no.	125	351	321	—
Vickers pyramid no.	—	341	314	771
static yield stress:				
(1) test machine	18 (*a*)	74 (*b*)	69 (*b*)	—
(2) ball method	29 (1)	85 (1)	70 (1), 85 (2), 79 (3)	127 (2)
dynamic yield stress by ball method	59 (1), 59 (3)	100 (1)	81 (1), 86 (2), 81 (3)	120 (1), 120 (4)
ratio of dynamic to static yield stress (ball method)	2·04	1·18	1·16, 1·01, 1·03	0·96

Notes. All yield stresses are in tons/in.^{2}. (*a*) denotes a compression test, (*b*) a tensile test. (1), (2), (3) and (4) denote that the tests were made with balls of diameters $\frac{1}{4}$, $\frac{1}{8}$, $\frac{3}{8}$ and $\frac{3}{16}$ in. respectively.

6. DISCUSSION

Table 6 shows two points of interest in connexion with the results of the present experiments. In the first place, the value of the static yield stress determined by the ball method is significantly higher than the value obtained by an ordinary tensile or compression test; and in the second place, there is a striking increase in the ratio of the dynamic and static yield stresses, found by the ball method, as the hardness number decreases.

Considering the first point, one possible explanation of the discrepancy is that the factor 0·63 in the equation $p_y = 0{\cdot}63p'_{\text{min.}}$ of p. 424–5, is in error on account of wrong assumption of the values of the elastic constants. Calculation shows that the factor depends only on the value of Poisson's ratio, σ, and that its values corresponding to the extreme values of σ for steel, namely 0·25 and 0·3, are 0·64 and 0·62 respectively. This difference is too small to explain the discrepancy in experimental results.

Another possibility of explaining the difference lies in the approximate nature of the Guest-Mohr criterion, which forms the basis of the method which gives the factor 0·63. At the same time, it should be pointed out that in the present case, because of the symmetry about O_z of figure 1, the equation $p_y = 0{\cdot}63\,p'_{\text{min.}}$ is again obtained if we use the criterion of maximum shear strain energy of von Mises. According to this criterion, plastic flow occurs when

$$\tau_1^2+\tau_2^2+\tau_3^2 = \tfrac{1}{2}p_y^2, \tag{6·1}$$

where τ_1, τ_2, τ_3 are the three principal shear stresses, given in terms of the principal stresses σ_1, σ_2 and σ_3 by

$$\tau_1 = \tfrac{1}{2}(\sigma_2-\sigma_3), \quad \tau_2 = \tfrac{1}{2}(\sigma_3-\sigma_1), \quad \tau_3 = \tfrac{1}{2}(\sigma_1-\sigma_2).$$

It can be shown that the expression on the left-hand side of equation (6·1) reaches its maximum on the z-axis, and for points on this axis, on account of symmetry,

$$X_x = Y_y, \quad Y_z = 0,$$

$$\sigma_1 = X_x = Y_y = \sigma_2, \quad \sigma_3 = Z_z,$$

so that

$$|\tau_1| = \tfrac{1}{2}(Z_z-X_x) = \tau_2; \quad \tau_3 = 0.$$

From equation (6·1), failure will therefore begin when and where the maximum value of

$$2(\tau_1^2+\tau_2^2+\tau_3^2) = (Z_z-X_x)^2 = (Z_z-Y_y)^2 = (\sigma_2-\sigma_3)^2$$

becomes equal to p_y^2. The criterion of maximum shear strain energy thus leads to the same result as the Guest-Mohr criterion of maximum difference of principal stresses, and therefore leads to the equation $p_y = 0·63p'_{\text{min.}}$.

A third factor, which may account for the discrepancy in question, is the difference in work-hardening between a tensile or compression test on the one hand, and the ball test on the other.

It may well be that all three factors contribute to the difference under discussion and, in this connexion, the results given in table III of Taylor's paper (1946), and the observation due to Mrs Tipper, that the yield-stress given by the ball method is much nearer to the 'ultimate stress', are of interest.

Turning next to the values of the yield stress obtained by the ball method under static and dynamic conditions, it is worth while pointing out that, since the stress distributions in the statical and dynamical experiments are identical (see p. 419), the ratio of the dynamical to the statical yield stresses, as determined by the present method, is independent of the criterion of failure; the ratio is therefore a real property of the material which is being tested. This is brought out very clearly by equation (II) of Taylor's paper,

$$p_{yd}/P_{ys} = 40·4\,h_{\text{min.}}^{\frac{1}{6}}\,R^{\frac{2}{3}}/M^{\frac{1}{3}} \tag{6·2)*$$

in the present notation; here the static load $P_{\text{min.}}$ is supposed to be due to gravity acting on a mass M g. ($P_{\text{min.}} = Mg$) and the dynamic load by a fall through a vertical distance $h_{\text{min.}}$ cm.

The general trend of the present results for different types of steel is in agreement with the results obtained by other workers. The curve given in figure 3 of Taylor's paper is interesting in this connexion. This curve shows the relationship between p_{yd}/p_{ys}, determined by the projectile method at the Road Research Laboratory, and the static yield stress obtained in a tensile test; the figure also shows values obtained

* This equation can be derived from equations (3·2) and (3·3) above.

by the ball method. Both sets of results show a decrease of p_{yd}/p_{ys} as static yield stress decreases, but the value of p_{yd}/p_{ys} given by the ball method for a given static yield stress is significantly lower than the value given by the projectile method. The difference is probably due to the fact mentioned earlier, that the ball method really tests only a very limited region of the specimen near the surface, whilst the projectile method gives a result averaged over the whole specimen, or, expressed differently, small local variations, due to the change in the properties of the layer caused by polishing can lead to considerable variations in the results given by the ball method, whereas effects of this kind are negligible in the projectile method.

It is a pleasure to acknowledge the advice and help given by Sir Geoffrey Taylor throughout the course of the work described in this paper.

The work was carried out for the Ministry of Supply to whom I am indebted for permission to publish.

REFERENCES

Coker, E. G. & Filon, L. N. G. 1931 *Photoelasticity*, p. 357, para. 4·23. Camb. Univ. Press.
Fry, A. 1921 *Krupp. Mh.* (July issue) (cf. Nadai, p. 91).
Fuchs, S. 1913 *Phys. Z.* **14**, 1282.
Hertz, H. 1881 *J. reine angew. Math.* **92**, 156. Reprinted in an English translation in Hertz's 'Miscellaneous Papers'. London: Macmillan.
Huber, M. T. 1904 *Ann. Phys., Lpz.*, **14**, 153.
Love, A. E. H. 1934 *Mathematical theory of elasticity*, paras. 138–140. Camb. Univ. Press.
Morton, W. B. & Close, L. J. 1922 *Phil. Mag.* **43**, 320.
Nadai, A. 1931 *Plasticity*, p. 91. New York: McGraw Hill.
O'Neill, H. 1934 *The hardness of metals and its measurement*, ch. I. London Chapman & Hall.
Southwell, R. V. 1936 *Theory of elasticity*, p. 140. Oxford: Clarendon Press.
Tabor, D. 1948 *Proc. Roy. Soc.* A, **192**, 247.
Taylor, Sir G. 1946 *J. Instn Civil Engrs*, **26**, 486.
Thomas, H. R. & Hoersch, V. A. 1930 *Bull. Univ. Ill. Engng Exp. Sta.* no. 212.

The magnetic field within the earth

By E. C. Bullard, F.R.S., *University of Toronto*

(*Received* 23 *November* 1948)

The paper discusses the magnetic effects of internal motions in the core of the earth. It is shown that tidal friction, fluctuations in the rate of rotation, nutation, and the variation of latitude have negligible magnetic effects. Precession is also ineffective if Poincaré's theorem on the precession of a liquid sphere in a rigid shell is applicable to the earth.

Thermal convection is shown to be likely to occur in the core. The conservation of angular momentum will require it to be associated with a radial gradient of angular velocity which will have a large magnetic effect. Its interaction with the dipole field can produce a toroidal field which is many times as intense as the dipole field.

The convective and rotational motions can interact with the dipole and toroidal field in a way that tends to reproduce the dipole field. The complete theory has not been worked out, but it seems likely that the interaction is strong enough to maintain the field. The whole process resembles that occurring in a self-exciting dynamo.

The existence of a strong field in the core removes the difficulties previously found in the theory of the secular variation.

1. INTRODUCTION

Many years ago Larmor (1919) suggested that the magnetic field of the earth might be maintained in a way analogous to that of a dynamo. He supposed that motions in the interior interact with the field to produce the electric currents necessary to maintain the field. In the 30 years since Larmor made this suggestion knowledge of the earth's interior has advanced, and it is now certain that there is a central liquid core having a radius of about half that of the earth. It is probable that this core is composed of molten iron and can thus provide the conducting medium required by Larmor.

To establish such a theory of the main field it is necessary to show that the motion produced by known agencies is of a kind that can maintain a field. This is the main problem of the present paper.

A clue to what is required is provided by the secular variation. A simple and natural explanation of this can be given if it is assumed to be due to electromagnetic induction by motions in the outer part of the core (Elsasser 1946; Bullard 1948, referred to hereafter as I). It is shown in I that such an explanation is only possible if the field in the core is substantially greater than that which would be obtained by applying the inverse cube law to the field observed at the surface. This makes it likely that a large field in the interior of the core is a necessary part of the mechanism that maintains the external field.

It has been pointed out by Elsasser (1947) that a large field can be produced in the core by the relative rotation of its parts. This is discussed quantitatively in § 2. In §§ 3, 4 and 5 possible astronomical causes of such a motion are discussed. The results are entirely negative, all the apparently adequate astronomical causes proving incapable of maintaining a sufficient differential rotation in the presence of the electromagnetic forces.

In § 6 the possibility of thermal convection in the core is examined. Convection will certainly occur if the temperature gradient required to transport the heat generated by radioactivity exceeds the adiabatic gradient. The rate of generation of

Library.

heat required for this is about 1 % of that occurring in the surface rocks and is of the same order as that in iron meteorites. It is found that the Corioli forces associated with convection would produce a radial variation in angular velocity of the kind required to give a large field in the core by interaction with the dipole field.

In § 8 the further interactions of the convective motion and the field are discussed, and it is shown to be probable that they will reproduce the main field, and thus maintain it. The calculations needed to render the argument complete are outlined.

2. THE CONDITIONS FOR A TOROIDAL FIELD

If a conducting sphere rotates in a magnetic field parallel to its axis, an electromotive force is produced directed at right angles to the axis of rotation. If the sphere rotates like a rigid body this produces no electric current. If, however, the angular velocity is not the same at all distances from the centre currents can flow. A magnetic field is then produced that runs along parallels of latitude, and circles the axis in opposite directions in the two hemispheres. This field is entirely confined within the sphere, and it is zero on the axis and in the equatorial plane. It is convenient to call it the toroidal field.

Elsasser (1946*a*, *b*, 1947) has discussed induction in a sphere with great generality and elegance. He does not, however, discuss particular models in detail. The purpose of this section is to take the simplest model of the core that will give the toroidal field and to determine the conditions necessary for the dipole field of the earth to give a toroidal field ten times as great as itself.

The model taken is one in which the outer and inner parts of the core each rotate like rigid bodies, but with slightly differing angular velocities. The details of the solution of this electromagnetic problem are somewhat complicated and are given in a separate paper (Bullard 1949, referred to as III).

Until its origin is completely understood the field within the core is uncertain. There is therefore a difficulty in knowing what field to take as the inducing field. Fortunately, the field deep in the core has only a slight effect, and it is possible to proceed by making extreme assumptions between which the truth must lie. The field outside the core is taken to be that of the centred dipole that best represents the field observed at the surface of the earth. Within the core we assume, first that the field is also that due to the centred dipole, and second that it is constant. Neither of these assumptions is plausible physically; they are intended to represent two extreme forms of field giving the same result at the surface of the core. One increases indefinitely as the centre is approached, the other is uniform throughout the core. Since both give the same result the uncertainty due to the lack of a theory of the main field is greatly reduced.

It may be shown (III, §§ 6 and 7) that if a sphere rotates in a uniform field H_0, the toroidal field produced in colatitude θ at the boundary between the rotating inner part of the core and the conducting shell is*

$$H_\phi = \tfrac{2}{5}\pi\kappa b v H_0(1-b^5/a^5)\sin 2\theta, \tag{1}$$

---

* If H_0 is the field on the axis at radius b the same expression gives the toroidal field due to a dipole at the centre.

where b is the radius of the inner sphere, a the outer radius of the shell (that is, the radius of the core, 3473 km.), v is the maximum relative velocity of the sphere and the shell, and κ is the electrical conductivity of both. H_0 will be taken as 4 gauss, the value of the dipole field at the surface of the core. For the conductivity we take 3×10^{-6} e.m.u. (3000 ohm^{-1} cm.$^{-1}$); this is the mean of the values (5000 and 1000 ohm^{-1} cm.$^{-1}$) deduced by different methods by Elsasser and the author, and is probably uncertain by a factor of about 3.

To remove the difficulties in the theory of the secular variation H_ϕ must be substantially greater than H_0, but its exact value is of no great importance. To obtain an order of magnitude for the rate of rotation required we take $H_\phi(\text{max.})/H_0 = 10$. The relative velocity needed can then be calculated as a function of b from (1). The results, expressed both as an angular and as a linear velocity at the surface of the rotating sphere, are shown in figure 1. It will be seen that the minimum velocity of 0·013 cm./sec. occurs when the conducting shell has a thickness of about 1000 km. The velocity is very insensitive to variations in this thickness, and for values between 210 to 2600 km. the velocity is in the range 0·013 to 0·03 cm./sec.

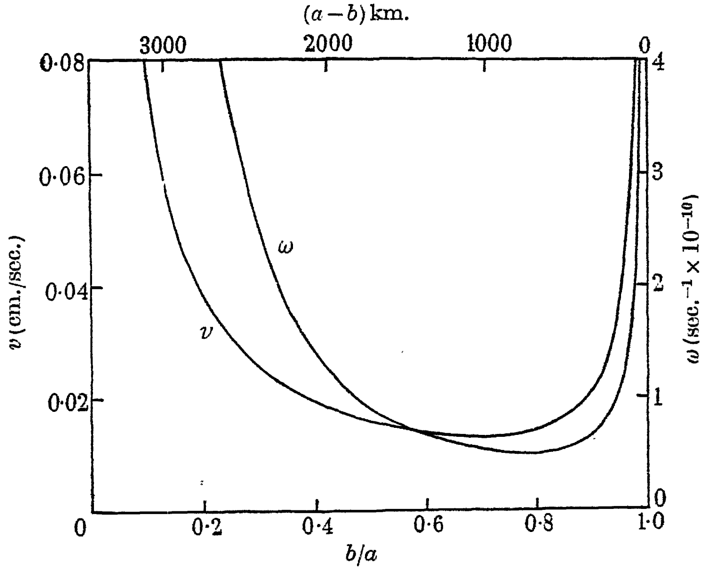


FIGURE 1. Linear and angular velocities necessary to produce a toroidal field of 40 gauss.

For a sphere of radius equal to that of the whole core rotating in an infinite conducting medium the peripheral velocity needed to give $H_\phi/H_0 = 10$ would be 0·0076 cm./sec., which is only about three times less than the velocities found in the last paragraph. Thus

$$H_\phi(\text{max.}) = 2\pi\kappa v a H_0/15 \tag{2}$$

may conveniently be taken as a rough approximation to H_ϕ except when $(a-b)$ is very small or very large. For shells thinner than 200 km. the velocity required increases rapidly and is given approximately by

$$v = 5{\cdot}3/(a-b),$$

where v is in cm./sec. and $(a-b)$ in km.

The conditions for the production of a toroidal field of 40 gauss are therefore quite definite. The inner part of the core must rotate relative to the outer. If the outer part is more than 200 km. thick a relative angular velocity of less than 10^{-10} sec.$^{-1}$ and a relative linear velocity of less than 0·03 cm./sec. will suffice. If the shell is thinner than 200 km. the velocity must be increased and reaches 1 cm./sec. for a thickness of about 5 km. The upper limit to the thickness of the shell will not concern us, as the difficulty is to get a thick enough shell rather than to avoid too thick a one. In making these calculations it has been assumed that the change from the shell to the inner part of the core is discontinuous. This is an analytical convenience, but it is certain that the substitution of a gradual change in velocity with radius would not affect the results in any important way. A gradual change is, of course, to be expected in a liquid.

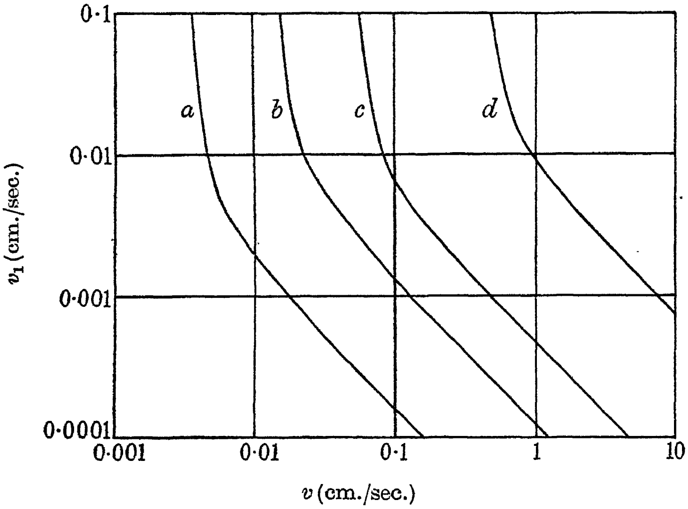


FIGURE 2. Velocities necessary to account for the secular variation. Curve a, cylinder, $a_1 = 400$ km.; curve c, sphere, $a_1 = 400$ km.; curve b, cylinder, $a_1 = 200$ km.; curve d, sphere, $a_1 = 200$ km.

The velocities necessary to produce the observed secular variation can be more closely investigated. It has been shown in I that the secular variation in South Africa during the last 100 years requires a dipole of moment $M = 2{\cdot}0 \times 10^{24}$ gauss cm.^{3} in latitude 25° S., longitude 20° E. or, alternatively, a long line of dipoles stretching from north-east to south-west through this point and having a moment of $3{\cdot}4 \times 10^{15}$ gauss cm.^{3}/cm. of its length. If the dipole is to be produced by the rotation of a sphere of radius a_1 with a peripheral velocity v_1, in a field H_ϕ, then

$$M = H_\phi a_1^3 f(\rho),$$

where $\rho^2 = 4\pi\kappa v_1 a_1$ and $f(\rho)$ is the function given in table 4 and figure 1 of III. If H_ϕ is taken as $(2/15)\,\pi v a H_0 \sin 2\Phi$, where Φ is the geomagnetic latitude, v_1 may be calculated for any assumed values of v and a_1. Similar relations hold for a cylinder. Typical results are shown in figures 2a and 2b. As would be expected a large v can be combined with a small v_1 or both may be of the same order of magnitude. There is, however, a lower limit below which v must not fall. For a sphere of radius 400 km.

the limit is 0·05 cm./sec., for a cylinder it is 0·0032 cm./sec. If v and v_1 are of the same order of magnitude both must be about 0·06 cm./sec. for a spherical eddy or 0·005 cm./sec. for a cylindrical eddy each 400 km. in radius. These figures may be somewhat underestimated, as no allowance has been made for the skin effect in the changing field of the secular variation (see I, § 8).

Velocities of the same order of magnitude will account for the widespread anomalies that constitute the 'non-dipole' part of the main field, for this field is of the same general magnitude as the secular variation summed over a hundred years.

When an attempt is made to fit the direction as well as the magnitude of the secular variation and non-dipole fields difficulties are found; these are discussed in §§ 7 and 8 following the development of the theory of the rotation of the core.

3. TIDAL FRICTION

The most obvious cause of a rotation of the inside of the core relative to the outer part is the slowing down of the whole earth by tidal friction. It might be expected that the slowing of the core would lag behind that of the solid part, and that the inner part of the core would, therefore, rotate more quickly than the rest. A boundary layer of the liquid core would be expected to cling to the outer solid part of the earth and to rotate with it. The theory of this process has been worked out in detail by Bondi & Lyttleton (1948) on the assumption that viscosity is the only force that has to be taken into account. They find that there is a boundary layer of thickness $\sqrt{(\nu/\Omega\,|\cos\theta\,|)}$, where ν is the kinematic viscosity, Ω is the angular velocity of rotation of the earth ($7{\cdot}29\times10^{-5}$ radian/sec.) and θ is the colatitude. The part of the core inside the boundary layer rotates with an angular velocity which varies radially. The maximum angular velocity occurs on the axis and is $\dot{\Omega}a/\sqrt{(\nu\Omega)}$ above that of the outer part of the earth, where $\dot{\Omega}$ is the secular deceleration of the earth ($\dot{\Omega} = 2{\cdot}2\times10^{-22}$ sec.$^{-2}$) and a is the radius of the core. The maximum linear velocity occurs at a radius $a/\sqrt{2}$ and is $\dot{\Omega}a^2/4\sqrt{(\nu\Omega)}$ above that for a rigid body. At the junction with the boundary layer the velocity falls to that characteristic of a rigid body rotating with the outer part of the earth. There is also a much slower motion towards the equatorial plane and away from the axis with a velocity of about $\dot{\Omega}a/\Omega$. In the boundary layer there is a drift from equator to poles with a velocity of about $\dot{\Omega}a^2/\sqrt{(\nu\Omega)}$. If we take* $\nu = 10^{-3}$ (I, § 9) these quantities become:

thickness of boundary layer	= 4 cm.
maximum angular velocity	= 3×10^{-10} sec.$^{-1}$
maximum linear velocity	= 0·025 cm./sec.
drift towards equator	= $1{\cdot}0\times10^{-9}$ cm./sec.
drift to poles in boundary layer	= 0·10 cm./sec.

The boundary layer in which the poleward motion occurs is too thin to have any perceptible magnetic effect. If its thickness is increased, as it might be if the laminar motion were unstable, the mean velocity in it would be reduced approximately in proportion, and would be only 4×10^{-7} cm./sec. for a thickness of 10 km. The velocity towards the equatorial plane is so small as only to produce a motion of 100 km. since the Cambrian. It seems, therefore, that the motion in meridian planes can be ignored.

---

* Bondi & Lyttleton take $\nu = 10^6$, a value which they obtain by extrapolating from laboratory conditions allowing for the effect of pressure, but neglecting the effect of temperature. Such a high value seems to me highly improbable.

The motion of rotation is at first sight hopeful, as its velocity is of the required order of magnitude and its possible magnetic effects have been suggested by several workers in this field. The hopeful appearance is, however, deceptive as may be shown by considering the energy available. The rate of loss of energy by the core is $I\Omega\dot{\Omega}$, where I is the moment of inertia of the core. Putting $I = 8{\cdot}6\times 10^{43}$ g.cm.², this gives $1{\cdot}4\times 10^{18}$ erg/sec. The energy in the magnetic field is shown in III, § 6 to be about $0{\cdot}13a^3H_\phi^2$(max.). With $H_\phi = 40$ gauss this gives 9×10^{27} erg. This energy will need to be replenished in a time comparable with the time in which the field would decay to $1/e$ in the absence of rotation. This time is about $4a^2\kappa/\pi$ or $4{\cdot}6\times 10^{11}$ sec. (14,000 yr.), which gives 2×10^{16} erg/sec. as the energy needed to maintain the field. If, therefore, 1·5 % of the energy lost by the core could be used to maintain the toroidal field we should have discovered an adequate source of power. This is, however, impossible, as all but a minute fraction of the energy lost by the core is transferred to the solid part of the earth. For consider a state in which the core is rotating faster than the outer part of the earth, and ignore for the moment the action of the moon. The rate of loss of energy by the core, $-\dot{E}_1$, is $-I\Omega\dot{\Omega}$, whilst the rate of loss of energy by the whole system, $-\dot{E}$, is

$$-\dot{E} = I_1\Omega_1\dot{\Omega}_1 + I\Omega\dot{\Omega},$$

where I_1 and Ω_1 are the moment of inertia and angular velocity of the solid part of the earth. This energy is that dissipated in the core. The conservation of angular momentum requires

$$I_1\dot{\Omega}_1 + I\dot{\Omega} = 0.$$

Thus

$$\dot{E} = I\dot{\Omega}(\Omega_1 - \Omega) = \dot{E}_1(\Omega_1 - \Omega)/\Omega,$$

since $(\Omega_1 - \Omega)/\Omega$ is a fraction of the order of 10^{-5} the proportion of the energy dissipated is much too small to be of any importance. The lunar forces exerted through the tides on the outer part of the earth will not affect this argument; they merely serve to absorb the energy transferred to the outer part of the earth.

The matter may be looked at from another point of view. The couple G tending to prevent relative rotation between the core and the outer part of the earth is (III, § 9)

$$G = \tfrac{4}{15}b^3H_0H_\phi(\text{max.}),$$

if $H_0 = 4$ gauss, $H = 40$ gauss, and $b = 3{\cdot}4\times 10^8$ cm. then $G = 1{\cdot}8\times 10^{27}$ dyne cm. The couple acting on the core is $I\dot{\Omega}$, which is $1{\cdot}9\times 10^{22}$ dyne cm. and is smaller than G by a factor of about 10^5. Thus the electromagnetic forces compel the rotation of the core to follow the rest of the earth with only a very small lag, and the forces of tidal friction could only produce about 4×10^{-4} gauss of toroidal field.

Elsasser (1947, p. 831) has suggested that the inner part of the core may rotate slower than the outer part owing to the electromagnetic dissipation of tidal energy in it. The problem has not been worked out in detail, but it seems improbable that there can be a perceptible effect of this sort. It is shown in III, § 11 that a magnetic field extending throughout the core takes a time of the order of 14,000 years to establish itself. The periods of the tides are so short compared with this that only very small currents will be produced. In the somewhat analogous problem of the oscillating sphere treated in III, § 10 this reduced the field by a factor equal to the square root of the ratio of the period to the time of relaxation of the field. For a semi-diurnal tide this factor is 1/3000. The maximum velocity of the equilibrium lunar semi-diurnal tide at the surface of the core is 2×10^{-3} cm./sec.; there is, therefore, no large velocity to offset the unfavourable period. If induction by the tidal motion cannot produce an appreciable field the energy dissipated will be small, and the field produced indirectly, by the slowing of the core, will be smaller still. Viscous dissipation of tidal energy in the core is easily shown to be negligible.

4. FLUCTUATIONS IN THE EARTH'S ROTATION

It has long been known that the angular velocity of the earth is not constant. The irregularities in its rate of rotation appear as apparent errors in the positions of the moon and planets. From a comparison of the results for different bodies it has been shown by Spencer Jones (1939) and de Sitter (1927) that the variations are almost wholly due to changes in the moment of inertia of the earth. It is noteworthy that these fluctuations are the only phenomenon with an origin inside the earth that has a time scale comparable with that of the secular variation; a connexion between the two is, therefore, not unreasonable.

Reliable observations extend from 1640 to the present day. The largest known sudden change occurred in 1897, when the day increased by about 0·0034 sec., corresponding to a decrease of about 3×10^{-12} sec.$^{-1}$ in the angular velocity of the earth. If the inner part of the core retained the angular velocity that it had before the change, a toroidal field of 3 gauss could be produced.

This is too small to be of much interest, but in view of the uncertainty in the true value of κ it is desirable to consider the matter in rather more detail. Suppose that the earth has been rotating for a very long time with a difference of angular velocity ω between the inner and outer parts of the core. If the couple maintaining the relative angular velocity is removed there will (III, § 9) be a decelerating electromagnetic couple

$$G = -\tfrac{4}{15}a^3 H_0 H_\phi \doteqdot -8\pi\kappa a^5 H_0^2 \omega/225; \qquad (4)$$

H_ϕ will decay with a time constant $\tau_1 = 4\kappa b^2/\pi$, which is about 14,000 years. For times short compared to this the couple will be given by (4) and

$$I\dot{\omega} = -8\pi\kappa a^5 H_0^2 \omega/225.$$

The angular velocity therefore decays exponentially with a time constant

$$\tau_2 = 225 I/8\pi\kappa a^5 H_0^2 = 15D/\kappa H_0^2,$$

where D is the density. Putting $D = 10{\cdot}7$ g./cm., $\kappa = 3 \times 10^{-6}$ e.m.u. and $H_0 = 4$ gauss, we get

$$\tau_2 = 39 \text{ days}.$$

The initial conditions assumed in this problem are, however, not what is required for considering the effects of a sudden change in the earth's rotation. If the earth has been rotating for a long time as a rigid body there will be no toroidal field. After a sudden change ω in the angular velocity of the outer part, the toroidal field will grow with a time constant τ_1 provided the relative angular velocity is maintained. In fact ω will be reduced by electromagnetic forces. A detailed treatment of the interaction between the growing field and the decreasing rotation would be complicated, but limits to the time τ required to reduce the angular velocity to $1/e$ are easily obtained by noting that for $0 < t < \tau$ the toroidal field is less than if ω had been maintained at its initial value, but greater than if it had its value at $t = \tau$. The result is

$$\tau = \beta\sqrt{(\tau_1\tau_2)}, \quad \text{where} \quad 1 < \beta < 2;$$

if $\tau_1 = 14{,}000$ years and $\tau_2 = 39$ days this gives $\tau \doteqdot 39$ years. The field can never greatly exceed its value at this time, which is less than $\beta\sqrt{(\tau_2/\tau_1)}$ times the value it would reach if ω were maintained. Since $\sqrt{(\tau_2/\tau_1)}$ is only 0·0028, and even the equilibrium field is unpromisingly low, it is impossible that sudden changes in rotation of the order observed could have any appreciable effect on the magnetic field.

Even the large change in moment of inertia due to the melting of the ice at the close of the ice age has only a moderate effect. If this change is represented very crudely as the sudden addition to the earth of a shell of water of thickness q, the material being previously at the poles, the change δI in the moment of inertia is

$$\delta I/I = 2(MR^2/I)\,Dq/D_0R,$$

where D is the density of the water, D_0 the mean density of the earth and M and R the mass and mean density of the earth. Putting $q = 10^4$ cm. and $I/MR^2 = 0{\cdot}33$ this gives $\delta I/I = 1{\cdot}7 \times 10^{-5}$, and the change in the day as 1·5 sec. This is about 500 times greater than the sudden changes occurring at present and could produce a toroidal field of a few gauss. As the value taken for q is moderate and the assumptions made are very crude, the matter is perhaps worthy of further attention. We shall not discuss it further here.

The result that a sphere rotating freely in a conducting medium in the presence of a magnetic field will have its angular velocity reduced to $1/e$ in a time τ_2 if it has a fully developed toroidal field, and in a time $\beta\sqrt{(\tau_1\tau_2)}$ if it has none, throws some light on the nature of the movements possible for a fluid in a magnetic field. If rotatory motions are to persist for a time long compared to τ_1 and τ_2 they must be driven by forces, and not merely maintained by inertia. For large bodies both τ_1 and τ_2 are very short compared to the time needed for viscosity to slow down rotation. The latter time is about $0{\cdot}05a^2/\nu$, which is 160 million years for a sphere of radius 100 km. and kinematic viscosity $\nu = 0{\cdot}001$. Thus the electromagnetic forces will be a much more potent agency than viscosity in the prevention of turbulence.

It is possible, but not very likely, that part of the variation in the earth's rate of rotation is periodic. In particular, attention has been given to the representation of the errors in the moon's longitude by Newcomb's 'great empirical term'

$$10''{\cdot}71 \sin(140^\circ{\cdot}0T + 240^\circ{\cdot}7),$$

where T is measured in centuries from A.D. 1900. This term could be produced by a fluctuation

$$6{\cdot}3 \times 10^{-10} \cos (140^\circ{\cdot}0 T + 240^\circ{\cdot}7)\ \text{radians/sec.}$$

in the angular velocity of the earth. The period τ_0 of this term is 257 years which is much less than τ_1, the natural period of decay of the current system (about 14,000 years). It is shown in § 10 of III that if the inner part of the core is unaffected by the varying speed of rotation the toroidal field will have an amplitude $0{\cdot}225\sqrt{(\tau_0/\tau_1)} = 0{\cdot}030$ time that which would be produced by a steady difference of $6{\cdot}3 \times 10^{-10}\,\text{sec.}^{-1}$ in angular velocity. From (2) above this gives a maximum toroidal field of 23 gauss. Since the period of 257 years is very large compared with the time of 39 days required for a fully developed toroidal field to reduce a difference of rotation to $1/e$, the inner part of the core will in fact follow closely the rotation of the outer part, and only a small fraction of the 23 gauss calculated above will be produced.

From these considerations it appears that neither the secular slowing of the earth's rotation, nor the sudden changes, nor the 'great empirical term' can have any important magnetic effects, in spite of the angular velocities being of the magnitude shown in § 2 to be necessary to produce a large toroidal field.

5. PRECESSION AND NUTATION

Poincaré has shown (Lamb 1932, p. 724) that the ellipticity of the earth is sufficient to ensure that the material of the core moves with the rest of the earth like a rigid body in a small motion of precession, even in the absence of viscosity. If this theorem is applicable to the earth, precession can have no magnetic effect. It has not been proved that the motion like a rigid body is stable, nor that it is possible for a precession of finite amplitude (the angle at the vertex of the cone swept out by the earth's axis is 47°). It is therefore not impossible that there may be internal motions associated with precession. If the inner part of the core failed to precess, a relative angular velocity of $3{\cdot}5 \times 10^{-5}\,\text{sec.}^{-1}$ would be produced. This is 10^5 to 10^6 times what is required to produce a large field in the core. It is therefore clear that Poincaré's theorem must hold with great exactness if precession is not to have a large magnetic effect. We do not treat the hydrodynamical problem further here, as Bondi & Lyttleton state that they will shortly discuss it. The relative angular velocity would be roughly at right angles to the earth's axis and the field produced of the type discussed in § 6 of III. The electromagnetic forces will tend to prevent relative motion and will probably, as in § 4, prevent the field rising above $\sqrt{(\tau_2/\tau_1)} = 0{\cdot}0028$ of the value it would attain if the full relative angular velocity of the precession were possible. In view of the large angular velocity available this argument does not prove the effect to be negligible, and everything hangs on the applicability of Poincaré's theorem.

Nutation, the secular change in the obliquity of the ecliptic and the variation of latitude may be shown to have negligible magnetic effects. Thus, with the possible exception of the precession, we have shown that astronomical causes are incapable of producing a field inside the core of the earth substantially greater than the dipole field. We now consider the effects of thermally induced motions.

6. Convection currents

It is improbable that the core is entirely free from radioactivity, and the heat produced by a very slight activity may cause convective motions with important magnetic effects.

Convection will certainly occur if the temperature gradient slightly exceeds the adiabatic gradient. The adiabatic gradient is $g\alpha T/c_p$, where α is the coefficient of thermal expansion, g is the acceleration due to gravity and c_p is the specific heat at constant pressure. In the outer part of the core g is about 900 cm./sec.^{2}, T may be taken as about 10,000° C (Jeffreys 1929, p. 139). α and c_p may be roughly estimated from their values at ordinary temperatures and pressures, using theoretical results for the extrapolation (Mott & Jones 1936). The results are 9×10^{-6}/° C and 0·18 cal./g.° C. These values give an adiabatic gradient of 1·1° C/km. The thermal conductivity may be estimated as 0·18 cal./° C cm.sec.; with this conductivity and a gradient equal to the adiabatic gradient, conduction could transport $3{\cdot}0 \times 10^{12}$ cal./sec. from the whole core. If the total heat generation exceeds this, convection must occur. Since the mass of the core is $1{\cdot}9 \times 10^{27}$ g. the corresponding generation of heat in each gram of the core is $1{\cdot}6 \times 10^{-15}$ cal./sec. This is under 1 % of the average rate of generation of heat in acid igneous rocks and is, therefore, not unreasonably high. It is, however, considerably higher than the rate of generation of heat in iron meteorites which may represent a material similar to that of the core.

Six meteorites examined by Paneth (1942) were found to contain an average of $0{\cdot}7 \times 10^{-8}$ g. U/g. and $4{\cdot}3 \times 10^{-8}$ g. Th/g. These would generate $0{\cdot}4 \times 10^{-15}$ cal./g.sec.,* or $0{\cdot}75 \times 10^{12}$ cal./sec. in the whole core, that is, about a quarter of the heat needed to maintain the adiabatic gradient against conduction. The data used are subject to such large errors that it is not certain that the gradient in a core composed of iron meteorites would really fall below the adiabatic. Even if it does it is still likely that convection currents will occur. This follows from an extension of von Zeipel's theorem due to Verhoogen (1948). He shows that a rotating ellipsoid in hydrostatic equilibrium cannot also be in thermal equilibrium, and must necessarily be subject to convection currents.

Convection currents may also be produced by asymmetries in the distribution of temperature. Further investigation is necessary to determine how rapid a motion will be produced by a given asymmetry.

It thus appears that the existence of convection currents in the core may reasonably be postulated, and that a temperature gradient of the order of the adiabatic is not unreasonable. To be on the safe side we shall assume one-tenth of the adiabatic gradient and show that if convection occurs it can produce a large toroidal field, and will probably explain the maintenance of the main field itself. The possible magnetic effects of convection have previously been discussed by Frenkel (1945), who suggested that it may be possible to base a theory of the main field on it, but who does not describe any specific mechanism. Slichter & Bullard (1947) have objected that convective motions on a large scale are unstable and would break up into a turbulent motion. This difficulty is removed by the arguments of § 4 which show that the magnetic field will exert a stabilizing influence on the motion.

---

* This figure is given incorrectly by Slichter & Bullard (1947).

The general effect of any convective motion is to interchange the material near the outside of the core with that inside. The conservation of angular momentum will then require that the inside of the core is continually accelerated and the outside retarded. In the presence of a dipole magnetic field a toroidal field will be produced and the accelerations will be opposed by the electromagnetic forces. The whole system constitutes a heat engine for converting the heat produced by radioactivity into the energy of a system of currents and a magnetic field. By the second law of thermodynamics the efficiency of such an engine cannot exceed $\delta T/T$, where δT is the difference of temperature over which it is working. We take this difference of temperature to be that produced by a gradient of one-tenth of the adiabatic gradient over a distance of the order of the radius of the core. Over 1000 km. one-tenth of the adiabatic gradient of 1·1° C/km. gives a temperature difference of 110° C. If $T = 10{,}000°$ C the efficiency cannot therefore exceed 1·1 %. The energy required to maintain a toroidal field of 40 gauss is about 2×10^{16} ergs/sec. or 5×10^{8} cal./sec. The efficiency required is therefore only 0·07 % or one-sixteenth of the ideal thermodynamic efficiency. In fact, there seems every reason why the process contemplated should give an efficiency not greatly below that for an ideal engine. Viscosity will dissipate an amount of energy that is negligible compared with that used in inducing the field, and eddy diffusion will probably be largely prevented by the field. Thermal conduction will be the main cause of irreversible changes, but if convection occurs at all it is likely to carry a large part of the heat. We conclude therefore that there is no difficulty in providing energy for convection currents in the core sufficient to produce a large magnetic effect.

In order to estimate roughly the radial velocities necessary to produce an appreciable toroidal magnetic field we consider the transport of momentum across the spherical surface that divides the core into two equal halves. This has a radius 0·79 time that of the core. If the core is rotating like a rigid body with an angular velocity Ω, the angular momenta M_1 and M_2 in the two halves are

$$M_1 = 0{\cdot}126 Ma^2\Omega, \quad M_2 = 0{\cdot}273 Ma^2\Omega,$$

where M and a are the mass and radius of the core. If the material of the two halves is interchanged in the absence of a magnetic field the conservation of angular momentum requires that the inside and outside halves should have angular velocities ω_1 and ω_2 after the change given by

$$0{\cdot}273 Ma^2\Omega = 0{\cdot}126 Ma^2\omega_1, \quad 0{\cdot}126 Ma^2\Omega = 0{\cdot}273 Ma^2\omega_2,$$

which gives $$\omega_1 - \Omega = 1{\cdot}17\Omega, \quad \Omega - \omega_2 = 0{\cdot}54\Omega, \quad \omega_1 - \omega_2 = 1{\cdot}71\Omega. \tag{5}$$

If the mean radial velocity over half the surface of radius $0{\cdot}79a$ is v_r inwards and over the other half v_r outwards, the time taken for the interchange is $0{\cdot}53a/v_r$. The rate of change of angular momentum of either half is

$$0{\cdot}126 Ma^2(\omega_2 - \Omega)\, v_r/0{\cdot}53a = 0{\cdot}28 Ma\Omega v_r.$$

This must be equal to the electromagnetic couple which is $\frac{4}{15}a_3 H_0 H_\phi$, whence, if $H_\phi = 40$ gauss,

$$v_r = a^2 H_0 H_\phi / 1{\cdot}0 M\Omega = 1{\cdot}4 \times 10^{-4}\,\text{cm./sec.} \tag{6}$$

This is the mean velocity.

The relative tangential velocity, v, of the two halves of the core corresponding to any given radial velocity can be found by substituting for H_ϕ in (2) from (6). The result is

$$v/v_r = 2{\cdot}4 M\Omega/\kappa a^3 H_0^2 = 160.$$

The rotation is, therefore, rapid compared with the radial motion.

The time to travel 1000 km. at a velocity v_r is $0{\cdot}7 \times 10^{12}$ sec. or 23,000 years. This gives plenty of time for the electromagnetic forces to redistribute the angular momentum before a complete circuit is completed.

If T_1 is the difference in temperature between the ascending and descending currents the rate of transport of heat across a surface of radius r is $2\pi r^2 D v_r c T_1$, where D is the density of the core and c its specific heat. If most of the heat is carried by convection, the temperature difference required to transport the $0{\cdot}75 \times 10^{12}$ cal./sec. assumed earlier in this section is 0·006° C. As the assumed heat is considerably more than is required to provide the energy for the field, this may be an overestimate. It is essential to the argument that this small difference of 0·006° C is *not* the temperature difference between which the engine works. It is the difference in temperature between the descending and ascending currents at a given level. The temperature difference occurring in the expression for the efficiency is the radial difference provided by the adiabatic gradient. Elsasser (1947) has not made this distinction, and has thus rejected radioactive heat as a factor in the problem. Frenkel (1945) makes a distinction similar to that made in this paper.

The mechanism described above decelerates the outer part of the core and tends to make it rotate more slowly than the rigid rocky mantle. Since the eddies causing the secular variation lie in the outer part of the core they may be expected to drift slowly westwards. This drift will cause the centres of rapid secular change to drift westwards with a velocity $(\Omega - \omega_2) R = 0{\cdot}32(\omega_1 - \omega_2) R$, where R is the radius of the earth. By (2) this velocity is about $3{\cdot}0 H_\phi R/\pi\kappa a^2 H_0$. For $H_\phi = 40$ gauss this gives 0·017 cm./sec. or 5·4 km./yr. A drift in the predicted direction is shown by Vestine's maps (Vestine, Laporte, Cooper, Lange & Hendrin 1947; Elsasser 1949); the observed velocity is about 7 km./yr. corresponding to $H_\phi = 50$ gauss and $\omega = 1{\cdot}4 \times 10^{-10}\,\text{sec.}^{-1}$. If the uncertainty in κ can be reduced, this method may give a reliable determination of H_ϕ and ω.

The boundary conditions for a viscous fluid require it to move with the solid with which it is in contact at the surface of the core. The work of Bondi & Lyttleton (1948) makes it likely that the transition layer will be so thin as to have no direct magnetic effect. The viscous forces in it will, however, exert a couple on the solid part of the earth. If this couple is large enough it may prevent the outer part of the core from rotating slower than the solid part of the earth and may thus invalidate the argument of the last paragraph. The viscosity and the thickness of the boundary layer are both so uncertain as to render a useful calculation impossible.

The ultimate fate of the $0{\cdot}75 \times 10^{12}$ cal./sec. that we have assumed to be generated in the core requires some consideration. It represents a small fraction of the whole flow of heat reaching the surface of the earth, which is about 7×10^{12} cal./sec., and its arrival at the surface would cause no difficulty. The difficulty is to devise a method of transporting it away from the core. Conduction cannot do this, as the whole of geological time is too short for an appreciable fraction of the heat generated in the core to be conducted to the surface. If convection currents exist in the outer part of the earth this difficulty is removed. The possibility of such currents has been discussed by many authors, but lies outside the scope of this paper.

If there are to be convection currents in the core the heat must leave it. If it did not leave, it would cause a rise in temperature. With the amounts of U and Th found by Paneth in meteorites the rise since the solidification of the earth $3{\cdot}3 \times 10^9$ years ago may be shown to be about 300° C. By the time all the U and Th have disintegrated this would rise to 1400° C. These are very moderate figures and show that the heat generated is not sufficient to have any rapid catastrophic effects. The time required to heat the whole core through the calculated difference in temperature between the ascending and descending currents is about 140,000 years. In this time the convection current would have travelled about 3600 km. or a distance of about one radius of the core.

The only feature of the convection currents that has been used in the argument of this section is the production of a difference in the angular velocity in the inner and outer parts of the core. Without a detailed theory of the motion it is difficult to go further. It is clear that the motion produced is not necessarily a rotation having the same angular velocity in all latitudes, and also that the magnetic effect of the radial motion must be considered. These factors make it likely that there will be some deviation from the simple picture of an easterly field in the northern hemisphere and a westerly one in the southern hemisphere. In the next section we show that the secular variation requires such complications, and in § 8 use them to outline a theory of the origin of the main field.

7. THE OBSERVED SECULAR VARIATION

Figure 3 shows a simplified picture of the secular variation of the vertical component of the field in 1922·5. This figure is derived from Vestine *et al.* (1947, figure 133*a*). It shows the line of no variation and the main centres of rapid change. The analysis in I of the region near South Africa suggests that this field resembles that which would be produced by a line of horizontal dipoles situated near the surface of the core and lying directly under the line of zero rate of change of vertical field. Calculations are in progress to test the accuracy of this approximation which seems likely to give a fair representation to the true field. Such a line of dipoles could be produced by a sinuous line of rotating matter with its axis along the line of dipoles. The long eddy assumed in I to explain the South African focus of rapid secular change would form part of this line. The field required in the core will depend on the radius assumed for the eddy and will be of the same general magnitude as those found in South Africa (e.g. 20 gauss for a cylinder of 200 km. radius; see I, p. 253). The magnitude now raises no difficulty, but the direction does. In its simplest form our theory gives opposite toroidal fields in the two hemispheres and only the dipole field at the equator. It is clear, however, from figure 3 that the lines of equal secular variation cross the equator without discontinuity. Also the field must have a component to the south-east under South Africa and to the south-west under South America. The first of these is inconsistent with a purely westerly field in the southern hemisphere. The secular variation thus requires a more complicated field for its explanation.

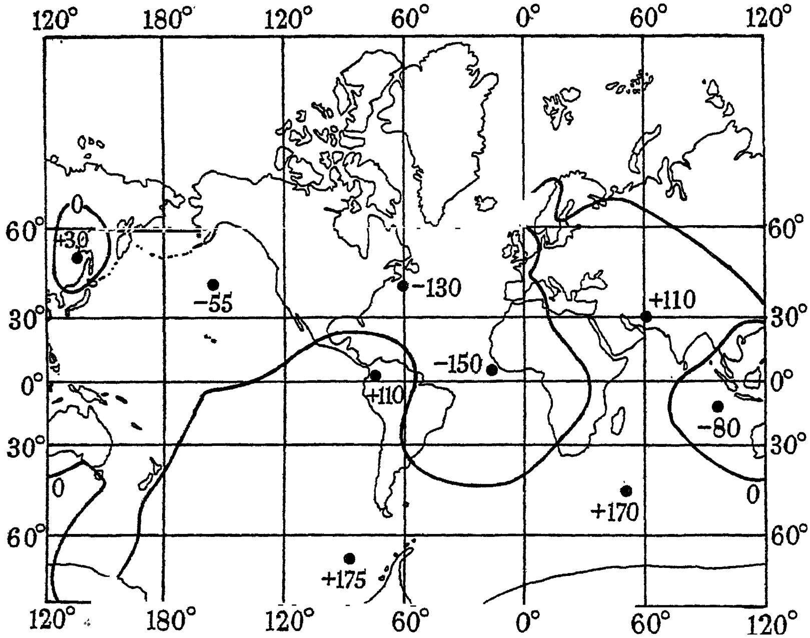


FIGURE 3. Secular change in the vertical component of the earth's magnetic field (γ/yr.) for 1922·5. The line of zero rate of change and all maxima of more than 30γ/yr. are shown.

8. THE CAUSE OF THE EARTH'S MAIN FIELD

The demonstration that it is possible in a very simple and natural way to produce a field in the interior of the earth much in excess of the dipole field leads to the reconsideration of Larmor's suggestion that the main field itself might be supported by induction. Elsasser (1947) has discussed the relation between field and motion in a very general way. He concludes that there do exist chains of relations in which a field H_1 interacts with a motion M_1 to produce a field H_2 which interacts with another component of the motion M_2 to reproduce H_1. The particular chain that he considers he rejects as 'so complicated and artificial that it would hardly seem convincing'. It suffers from the almost insuperable objection that it involves a rotation about an axis inclined to that of the earth.

The considerations of § 6 suggest that the only plausible motion in the core on a large scale is a radial convective motion, together with the rotation about the earth's axis that is necessary to conserve angular momentum. We now attempt to construct a system of motions of this type that will maintain a field. For this it is desirable to employ Elsasser's notation for vector fields. An electric or magnetic field or a current system satisfying Maxwell's equations can be derived from scalar functions ψ_n^{mc} and ψ_n^{ms}, where

$$\left.\begin{aligned}\psi_n^{mc} &= R(r)P_n^m(\theta)\cos m\phi,\\ \psi_n^{ms} &= R(r)\,P_n^m(\theta)\sin m\phi,\end{aligned}\right\} \tag{7}$$

here θ is the colatitude, ϕ the longitude and $R(r)$ a function of r (see III, §3). The fields derived from this are of two kinds called toroidal and poloidal. The toroidal field T_n^{mc} has components

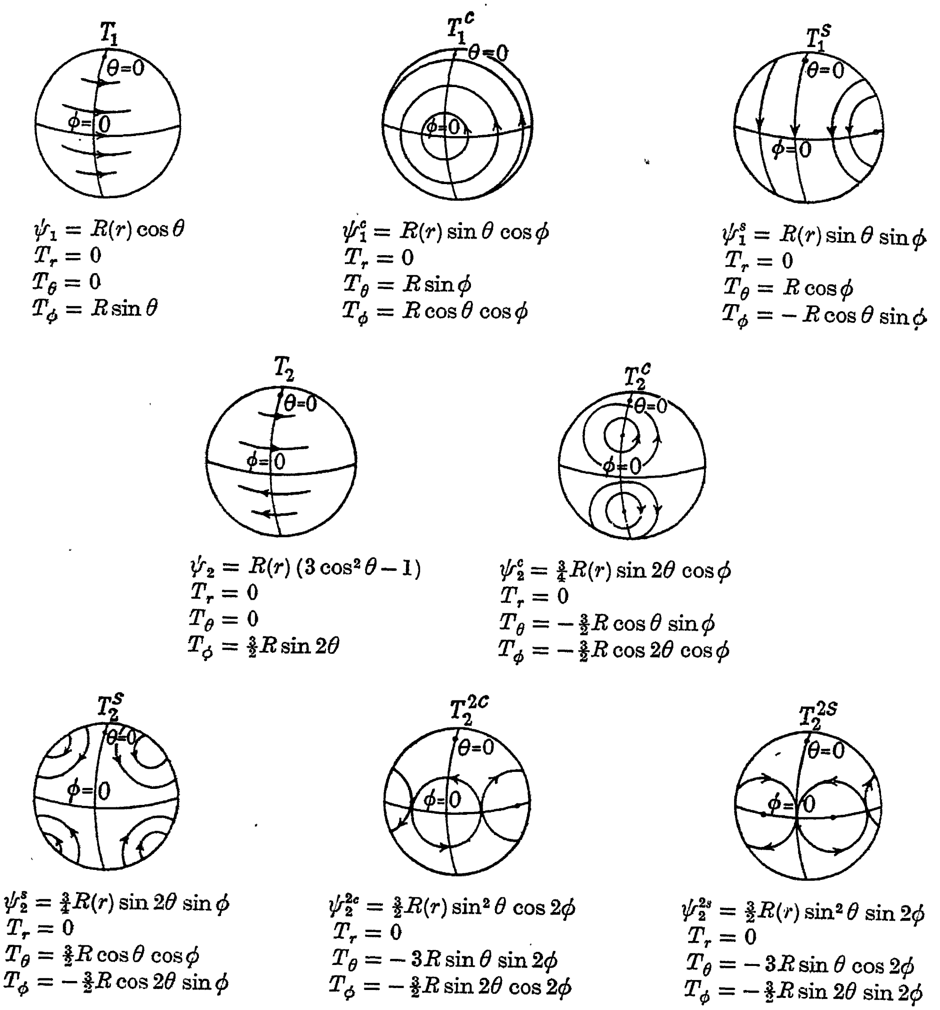
$$\begin{aligned}\text{radial component of } T_n^{mc} &= 0\\ \theta \text{ component of } T_n^{mc} &= (\sin\theta)^{-1}\partial\psi/\partial\phi\\ \phi \text{ component of } T_n^{mc} &= -\partial\psi/\partial\theta\end{aligned}$$

and similarly for T_n^{ms}. The corresponding poloidal vectors S_n^{mc} and S_n^{ms} are proportional to curl T_n^{mc} and curl T_n^{ms}. The general form of the fields for $n = 1$ and 2 is shown in figures 4*a* and *b*. The S fields are more difficult to visualize than the T's as they have a radial component. More than one sketch is given for some of the S fields to illustrate the behaviour both when confined within the sphere and when spreading outside it, the former apply to currents and the latter to magnetic fields.

The same scheme can be used to classify motions within a sphere. A system of convection currents in a stationary sphere might involve motions of the S_1, S_1^c, S_1^s, S_2^c, S_2^s S_2^{2c} or S_2^{2s} types. In a rotating sphere these would be accompanied by T_1 and perhaps by T_2 motions produced by the mechanism described in §6. There may also be S and T motions with higher n's, but these may probably be neglected in a preliminary discussion. It is possible, by methods devised by Elsasser, to draw up tables showing what interactions are possible between these motions and a given field (the tables given by Elsasser refer to complex vectors, but are easily adapted to the real vectors used here).

The field outside the earth is predominately S_1. It is natural therefore to start with this and to consider the network of possible interactions that stems from it. If this is done it is found that the combinations of motions $T_1 S_1$ and $T_1 S_2$ do not produce a closed chain returning to S_1, and therefore cannot maintain a field. The combinations $T_1 S_2^{2c}$, $T_1 S_2^c$ and $T_1 S_1^c$ do produce closed chains. Similar schemes would apply to the corresponding motions with s in place of c in the upper index.

The simplest of these schemes is that associated with an S_2^{2c} motion (figure 5 and table 1). This motion consists of two rising and two sinking currents spaced evenly around the equator. The S_1 field outside the core may be taken as that of a dipole at the centre of the earth. Inside the core its radial variation is not known, but it is probably approximately constant. This field interacts with the T_1 rotation (figure 6*a*); in this motion the inner part of the core rotates from west to east relative to the outer. This gives an S_2 current system (figure 6*b*); in this the current travels inwards along the equatorial plane and outwards at both poles. This current produces a T_2 magnetic field directed from west to east in the northern hemisphere and from east to west in the southern hemisphere (figure 6*c*). This



$\psi_1 = R(r)\cos\theta$
$T_r = 0$
$T_\theta = 0$
$T_\phi = R\sin\theta$

$\psi_1^c = R(r)\sin\theta\cos\phi$
$T_r = 0$
$T_\theta = R\sin\phi$
$T_\phi = R\cos\theta\cos\phi$

$\psi_1^s = R(r)\sin\theta\sin\phi$
$T_r = 0$
$T_\theta = R\cos\phi$
$T_\phi = -R\cos\theta\sin\phi$

$\psi_2 = R(r)\,(3\cos^2\theta - 1)$
$T_r = 0$
$T_\theta = 0$
$T_\phi = \frac{3}{2}R\sin 2\theta$

$\psi_2^c = \frac{3}{4}R(r)\sin 2\theta\cos\phi$
$T_r = 0$
$T_\theta = -\frac{3}{2}R\cos\theta\sin\phi$
$T_\phi = -\frac{3}{2}R\cos 2\theta\cos\phi$

$\psi_2^s = \frac{3}{4}R(r)\sin 2\theta\sin\phi$
$T_r = 0$
$T_\theta = \frac{3}{2}R\cos\theta\cos\phi$
$T_\phi = -\frac{3}{2}R\cos 2\theta\sin\phi$

$\psi_2^{2c} = \frac{3}{2}R(r)\sin^2\theta\cos 2\phi$
$T_r = 0$
$T_\theta = -3R\sin\theta\sin 2\phi$
$T_\phi = -\frac{3}{2}R\sin 2\theta\cos 2\phi$

$\psi_2^{2s} = \frac{3}{2}R(r)\sin^2\theta\sin 2\phi$
$T_r = 0$
$T_\theta = -3R\sin\theta\cos 2\phi$
$T_\phi = -\frac{3}{2}R\sin 2\theta\sin 2\phi$

FIGURE 4*a*. Classification of toroidal fields. $\phi = 0$ is to the front and $\phi = \frac{1}{2}\pi$ to the right of all the diagrams (T_1^c and T_2^{2s} have accidentally been drawn for negative R, the rest have R positive).

TABLE 1. INDUCTION BY T_1 AND S_2^{2c} MOTIONS

velocity	inducing field	current	induced field
T_1	S_1	S_2	T_2
S_2^{2c}	S_1	S_2^{2s}	T_2^{2s}
S_2^{2c}	T_2	S_2^{2c}	T_2^{2c}
S_2^{2c}	T_2^{2c}	S_2	T_2
S_2^{2c}	T_2^{2s}	T_1	S_1
T_1	T_2^{2c}	S_2^{2s}	T_2^{2s}
T_1	T_2^{2s}	S_2^{2c}	T_2^{2c}

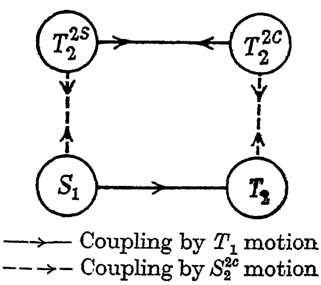
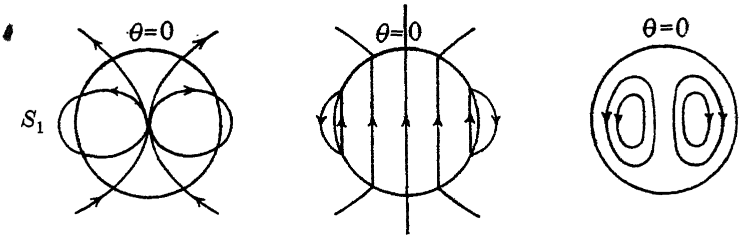


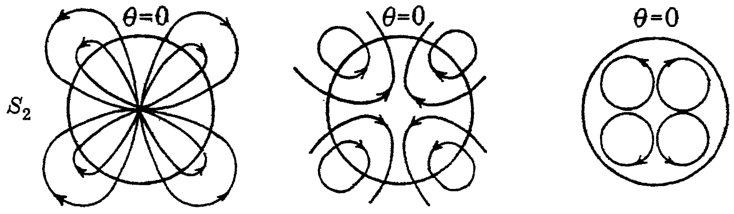
FIGURE 5. Relations between induced and inducing fields with T_1 and S_2^{2c} motions.



Dipole at centre

No pole at centre

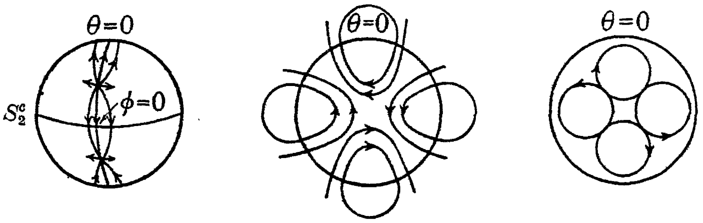
No external field
No pole at centre



Quadrupole pole at centre

No pole at centre

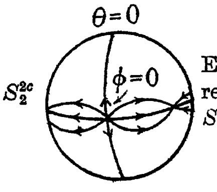
No external field and no pole



Projection on sphere

Section of previous diagram
in plane $\phi = 0$
No pole at centre

No pole at centre
No external field



Equatorial sections of S_2^{2c} fields resemble the above sections of S_2^c by the plane $\phi = 0$

Projection on sphere

Figure 4b. Classification of poloidal fields. The diagrams for S_1 and S_2 are sections in any meridian plane. These fields are symmetrical about the axis $\theta = 0$, and are the same in all meridian planes. The first diagram for S_1 represents a dipole at the centre, the second represents a uniformly magnetized sphere, and the third a field confined within the sphere. A field confined within the sphere with a dipole at the centre is also possible. The S_1^c and S_1^s are similar to S_1 but have their axes along $\theta = \frac{1}{2}\pi$, $\phi = 0$ and $\phi = \frac{1}{2}\pi$. S_2^s is similar to S_2^c turned through $\frac{1}{2}\pi$ about the axis $\theta = 0$. S_2^{2s} is similar to S_2^{2c} turned through $\frac{1}{4}\pi$ about the axis $\theta = 0$. The radial variation shown is diagrammatic only.

process has been thoroughly discussed in §§2 and 6, where it is shown that the T_2 field may be many times the dipole field. The only experimental evidence for its magnitude is the westward drift of the centres of secular variation. This suggests about 50 gauss.

In the S_2^{2c} motion material emerges in the neighbourhood of two diametrically opposite points on the equator at $\phi = 0$ and π and sinks near the points on the equator half-way between them. A projection of the motion on the outside of the core is shown in figure 6*c*, and an equatorial section in figure 6*d*. The T_2 field interacts with this motion to give an S_2^{2c} current. The origin of this current may be seen by considering the interaction of the northward component of the motion along the meridian $\phi = 0$ (figure 6*c*) with the westward T_2 field in the northern hemisphere.

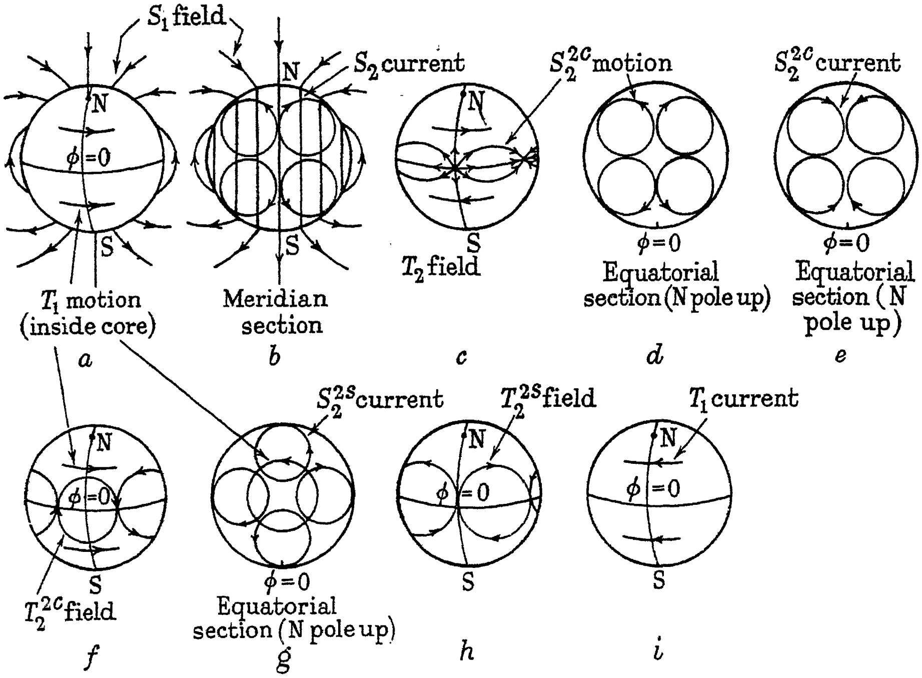


FIGURE 6. Motions, currents and fields required to maintain an external dipole field by the process of table 1 and figure 5.

This produces an electromotive force which has a component inwards. In the southern hemisphere the motion is southward and the field westward; the electromotive force is therefore also inwards. Similar arguments give an inward electric field at $\phi = \pi$ and an outward field at $\phi = \pm \frac{1}{2}\pi$. Completing the circuits of currents produced by these fields gives an S_2^{2c} current (figure 6*e*). This current cannot be geometrically similar to the S_2^{2c} motion that produces it. Their angular variation must be the same, but they will differ in a radial direction, that is, the $R(r)$ of (7) will differ. As the radial variation of neither is known this difference is not represented in figures 6 *d* and *e*.

The S_2^{2c} current gives a T_2^{2c} field (figure 6*f*). The interaction of this with the T_1 rotation gives a similar current system and field, with the centres of the circular lines of force still on the equator but no longer at $\phi = 0$ and $\frac{1}{2}\pi$ (figures 6 *g* and *h*); that is, some S_2^{2s} current and T_2^{2s} field are present as well as the S_2^{2c} and T_2^{2c}. This interaction is similar to that discussed in § 8 of III, except that the radial variation there assumed for the T_2^{2c} field is not the correct one for the present problem. The interaction will be strong and the T_2^{2s} field comparable with the T_2^{2c} one if $4\pi\kappa av$ is greater than 1, where v is the maximum velocity due to the T_1 motion. With a velocity of 0·03 cm./sec. $4\pi\kappa av$ is about 400. The T_2^{2c} and T_2^{2s} fields are therefore strongly coupled.

Finally, the T_2^{2s} field interacts with the S_2^{2c} motion to give a T_1 current (figure 6*i*) which reinforces the original S_1 field.

The directions of the fields and currents are given in figure 6, which demonstrates that the final field is in the correct direction to reinforce the initial field. This is the natural result of the chain and not the consequence of an arbitrary choice. The same result is obtained if the S_2^{2c} motion is reversed. The direction of the T_1 motion is fixed by the dynamical arguments of § 6. The argument does not, however, fix the direction of S_1; reinforcement could occur in either direction. If the earth were unmagnetized but possessed the T_1 and S_2^2 motions, a small field along the axis in either direction could start the process of regeneration, and cause the field to increase till the electromagnetic forces controlled the motions.

The external field produced by this process is that due to a dipole directed along the earth's axis. The internal field consists of three parts; an approximately uniform field of about 4 gauss associated with the external dipole, the large T_2 field discussed above and a T_2^2 field, probably of an intermediate magnitude. The T_2^2 field has a northerly or southerly direction near the equator, and is of the kind required to remove the difficulties of § 7 concerning the direction of the field producing the secular variation. A properly disposed T_2^2 field can have a component to the south-east in South Africa and to the south-west in South America.

The possibility of the process outlined depends essentially on whether the large gain in the step from the S_1 field to the T_2 is sufficient to balance possible losses in the steps from T_2 to T_2^{2c} and from T_2^{2s} back to S_1. It is important that this problem should be solved in detail. That is to say that a solution of Maxwell's equations should be found that is a combination of the S_1, T_2 and T_2^2 fields (and perhaps fields with $n > 2$), which satisfies the boundary conditions and which does not decay to zero.

Until such a solution is found we must rely on general arguments and on analogy with the problems solved in III. It is shown in III, § 2, that Maxwell's equations without the displacement current term may be written in a non-dimensional form so that the only parameter involved is $4\pi\kappa va$, where v is the maximum velocity occurring in the motion. The ratio of the induced to the inducing field in any interaction will be a function of this parameter only. If the interaction will only work in one direction, as in the production of a T_2 field from an S_1 field by a T_1 rotation, the induced field increases indefinitely as v is increased. If there is an interaction in both directions, as in the interaction of an S_1^c field with a T_1 rotation, the induced field may tend to a limit of the same order as the inducing field (see figure 1 of III). It seems likely that the two interactions with which we are concerned are of the latter type, and that if a large factor is not to be lost in each $4\pi\kappa va$ must not be much below unity. With the velocity of $1\cdot4 \times 10^{-4}$ cm./sec. found in § 6 $4\pi\kappa va$ is 1·8. There is thus a good prospect of only a moderate loss in the two steps, and that it may be counterbalanced by the large gain in the first step. The scheme is more complicated than Elsasser's, but the motions are simple and physically plausible. The complication is the inevitable result of Maxwell's equations.

The schemes with $T_1 S_2^c$ and $T_1 S_1^c$ motions are more complicated than that of table 1. They have several closed chains of interactions, and without a detailed solution it is not possible to know if the resultant of them all is in a direction to reinforce the dipole field. The three schemes are not mutually exclusive, and an actual system of convection currents may well involve all three. Only the S_1 and S_2 part of the motion, that is, the part that is symmetrical about the axis, does not enter into any of the schemes.

The scheme with T_1 and S_2^c motions automatically produces a dipole inclined to the earth's axis. Like that of table 1 and figure 5 it produces an external field with no second harmonics.

Cowling (1934) has shown that a motion that is purely in meridian planes cannot maintain a field, and Elsasser has shown that the same is true of a purely toroidal motion. Our schemes do not violate these theorems, as the motion is of a more complex type.

The theory, like all theories that ascribe a deep internal origin to the field, is inconsistent with the experiments of Hales & Gough (1947) and of Runcorn (Chapman 1948) on the variation of field with depth. It is not impossible that their results are due to local anomalies, and it is of great importance that the doubt should be removed by measurements at other places and especially at sea.

Before the theory can be regarded as giving a satisfactory explanation of the origin of the main field two things are necessary. First the convective motion occurring in a rotating sphere must be calculated and shown to be of a type capable of maintaining a field. The electromagnetic problem presented by this scheme must then be solved and the values of the fields calculated in terms of the velocities. The velocities found must be reasonable judged by the thermodynamic arguments of § 6.

The proposed calculations are formidable and the construction of a model is therefore worth consideration. Such a model might consist of a solid sphere of copper carrying inside it four copper cylinders which could be rotated and which would represent the four circulations of the convective motion. Outside the sphere would be a rotating copper shell representing the T_1 motion. The condition for similarity is that $\kappa a v$ should be the same in the model as in the earth. For copper κ is about 200 times greater than we have assumed for the earth. Thus if a is 20 cm., which is about the largest practicable size, v must be $0{\cdot}8 \times 10^5$ times that in the earth. For the convective motion this will be only a few cm./sec., and is easily attained. For a T_1 rotation of 0·1 cm./sec. in the earth, it would give about 10^4 cm./sec. or 5000 r.p.m. While this is not absolutely unattainable it is so high as to make the construction of a model a considerable undertaking and not justifiable till the calculations are further advanced.

9. CONCLUSION

The arguments of this paper are believed to demonstrate that a very moderate radioactivity of the material of the core will have profound effects on the magnetic field within the core. This field is larger and more complicated than appeared at first sight to be likely. There seems now no serious difficulty in accepting the induction theory of the origin of the secular variation and of the lower harmonics of the non-dipole field. The main field itself also finds a natural and unforced explanation, though the arguments used still fall short of a proof that the mechanism postulated will produce a field.

The explanation of the westward drift of the centres of secular variation is some confirmation of the general correctness of the ideas employed.

The only arbitrary factor in the calculation is the rate of generation of the heat in the core by radioactivity. This has been taken from studies of meteorites. It proves to be much greater than is necessary to supply the energy required, though somewhat smaller than is necessary to produce the adiabatic gradient.

In this paper I have differed in several matters from Professor Elsasser, and have to some extent built a theory with ideas which he has discarded for reasons that I consider inadequate. Without some further remark it might be thought that our views differed more than they do. In fact, the development of the ideas in this paper has been very greatly influenced by his work.

REFERENCES

Bondi, H. & Lyttleton, R. A. 1948 *Proc. Camb. Phil. Soc.* **44**, 345–359.

Bullard, E. C. 1948 *Mon. Not. R. Astr. Soc. Geophys. Suppl.* **5**, 248–257 (referred to as I).

Bullard, E. C. 1949 *Proc. Roy. Soc.* A. (Submitted for publication), (referred to as III).

Chapman, S. 1948 *Nature,* **161**, 462–464.

Cowling, T. G. 1934 *Mon. Not. R. Astr. Soc.* **94**, 39–48.

Elsasser, W. M. 1946*a* *Phys. Rev.* **69**, 106–116.

Elsasser, W. M. 1946*b* *Phys. Rev.* **70**, 202–212.

Elsasser, W. M. 1947 *Phys. Rev.* **72**, 821–833.

Elsasser, W. M. 1949 (in the Press).

Frenkel, J. 1945 *C.R. Acad. U.R.S.S. Sci.* **49**, 98–101.

Hales, A. L. & Gough, D. I. 1947 *Nature,* **160**, 746.

Jeffreys, H. 1929 *The earth.* Cambridge University Press.

Jones, H. Spencer 1939 *Mon. Not. R. Astr. Soc.* **99**, 541–558.

Lamb, H. 1932 *Hydrodynamics,* 6th ed. Cambridge University Press.

Larmor, J. 1919 *Rep. Brit. Ass.* pp. 159–160.

Mott, N. F. & Jones, H. 1936 *The theory of the properties of metals and alloys.* Oxford University Press.

Paneth, F. A. 1942 *Nature,* **149**, 235–238.

Sitter, W. de 1927 *Bull. astr. Insts. Netherlds,* **4**, 21–38.

Slichter, L. B. & Bullard, E. C. 1947 *Nature,* **160**, 157.

Verhoogen, J. 1948 *Trans. Amer. Geophys. Un.* **29**, 361–365.

Vestine, E. H., Laporte, L., Cooper, C., Lange, I. & Hendrin, W. C. 1947 *Description of the earth's main field and its secular change,* 1905–1945. Washington: Carnegie Institution.

Microscopy by reconstructed wave-fronts

BY D. GABOR, DR.-ING.

*Research Laboratory, British Thomson-Houston Company Ltd., Rugby**

(*Communicated by Sir Lawrence Bragg, F.R.S.—Received* 23 *August* 1948—
Revised 28 *December* 1948—*Read* 17 *February* 1949)

[Plates 15 to 17]

The subject of this paper is a new two-step method of optical imagery. In a first step the object is illuminated with a coherent monochromatic wave, and the diffraction pattern resulting from the interference of the coherent secondary wave issuing from the object with the strong, coherent background is recorded on a photographic plate. If the photographic plate, suitably processed, is replaced in the original position and illuminated with the coherent background alone, an image of the object will appear behind it, in the original position. It is shown that this process reconstructs the coherent secondary wave, together with an equally strong 'twin wave' which has the same amplitude, but opposite phase shifts relative to the background.

The illuminating wave itself can be used for producing the coherent background. The simplest case is illumination by a point source. In this case the two twin waves are shown to correspond to two 'twin objects', one of which is the original, while the other is its mirror image with respect to the illuminating centre. A physical aperture can be used as a point source, or the image of an aperture produced by a condenser system. If this system has aberrations, such as astigmatism or spherical aberration, the twin image will be no longer sharp but will appear blurred, as if viewed through a system with twice the aberrations of the condenser. In either case the correct image of the object can be effectively isolated from its twin, and separately observed. Three-dimensional objects can be reconstructed, as well as two-dimensional.

The wave used in the reconstruction need not be the original, it can be, for example, a light-optical imitation of the electron wave with which the diffraction diagram was taken. Thus it becomes possible to extend the idea of Sir Lawrence Bragg's 'X-ray microscope' to arbitrary objects, and use the new method for improvements in electron microscopy. The apparatus will consist of two parts, an electronic device in which a diffraction pattern is taken with electrons diverging from a fine focus, and an optical synthetizer, which imitates the essential data of the electronic device on a much enlarged scale.

The theory of the analysis-synthesis cycle is developed, with a discussion of the impurities arising in the reconstruction, and their avoidance. The limitations of the new method are due chiefly to the small intensities which are available in coherent beams, but it appears perfectly feasible to achieve a resolution limit of 1 Å, ultimately perhaps even better.

INTRODUCTION

The period of steady progress in the resolving power of electron microscopes which was started in 1931 by Knoll & Ruska came virtually to an end in 1946, when Hillier & Ramberg (1947) eliminated the astigmatism of their objective, and achieved a resolving power only insignificantly different from the theoretical limit. The barrier which stopped progress is of a technical nature, but formidable enough to prevent any really essential improvements along the direct line.

The theoretical limit of conventional electron microscopes is about 5 Å. It is determined by a compromise between diffraction and spherical aberration in electron objectives, and at the best compromise it is proportional to the fourth root of the aberration constant. Though several suggestions for correction have been put forward, they involve such technical difficulties that an improvement by a factor of 2 is the best that can be expected, even optimistically. One can never hope to achieve a resolving power ten times better than the present, which would require a correction of the spherical aberration to about 1 part in 10,000. Such precision can be realized with the technique of the optical workshop, but hardly ever with the means at the disposal of electron optics.

---

* Now at Imperial College, London, Electrical Engineering Department.

The new method is an attempt to get around the obstacle, instead of across it, by a two-step process, in which the analysis is carried out with electrons, the synthesis by light. The general idea of such a process was first suggested to the author by Sir Lawrence Bragg's 'X-ray microscope' (Bragg 1942; cf. also Boersch 1938). But Bragg's method, in which a lattice is reconstructed by diffraction from an X-ray diffraction pattern, can be applied only to a rather exceptional class of periodic structures. It is customary to explain this by saying that diffraction diagrams contain information on the intensities only, but not on the phases. The formulation is somewhat unlucky, as it suggests at once that since phases are unobservables, this state of affairs must be accepted. In fact, not only that part of the phase which is unobservable drops out of conventional diffraction patterns, but also the part which corresponds to geometrical and optical properties of the object, and which in principle could be determined by comparison with a standard reference wave. It was this consideration which led me finally to the new method.

In order to make the two-step method generally applicable, it had to be combined with a principle apparently not hitherto recognized. If a diffraction diagram of an object is taken with coherent illumination, and a coherent background is added to the diffracted wave, the photograph will contain the full information on the modifications which the illuminating wave has suffered in traversing the object, apart from an ambiguity of sign, which will be discussed later. Moreover, the object can be reconstructed from this diagram without calculation. One has only to remove the object, and to illuminate the photograph by the coherent background alone. The wave emerging from the photograph will contain as a component *a reconstruction of the original wave*, which appears to issue from the object. Conditions can be found in which the remainder can be sufficiently separated from the useful component to allow a true, or very nearly true, reconstruction of the original object.

This principle has been confirmed by numerous experiments. Some of the results are shown in figures 10 to 12 and explained in the last section of this paper.

In light optics a coherent background can be produced in many ways, but electron optics does not possess effective beam-splitting devices; thus the only expedient way is using the illuminating beam itself as the coherent background. This leads us to illumination by a coherent, divergent electron wave, illustrated in figure 1. It will be useful to explain this arrangement first, anticipating the principle of reconstruction which will be proved later.

The apparatus consists of two parts, the electronic analyzer and the optical synthetizer. The analyzer is similar to an electron shadow microscope (Boersch 1939), but with the important difference that it operates with coherent illumination, and under conditions in which the shadow microscope is useless, as the interference diagram has little likeness to the original. An electron gun, combined with a suitable aperture and electron lens system, produces a coherent illuminating beam, as nearly homocentric as possible. Exactly homocentric illumination is of course impossible, because of the unavoidable spherical aberration of electron lenses, but for simplicity we can talk of the narrow waist of the beam as of a 'point focus'. A small object is arranged some small distance before or behind the point focus, and a photographic plate at a comparatively large distance L. The divergence angle of the beam, γ_m, must be sufficient for the required resolution limit d_A, which is by Abbe's relation

$$d_A = \tfrac{1}{2}\lambda \sin \gamma_m.$$

The factor $\frac{1}{2}$ will be used in this paper to simplify the discussions, except in numerical calculations, where it will be replaced by the more accurate value 0·6.

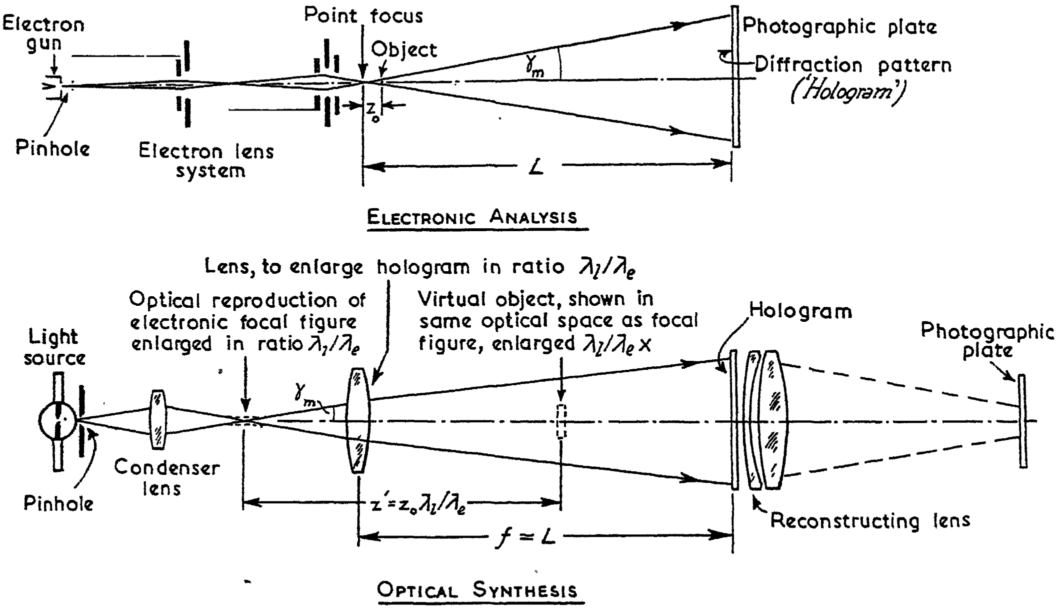

Electron gun
Pinhole
Electron lens system
Point focus
Object
z_0
γ_m
Photographic plate
Diffraction pattern ('Hologram')
L
ELECTRONIC ANALYSIS
Lens, to enlarge hologram in ratio λ_l/λ_e
Optical reproduction of electronic focal figure enlarged in ratio λ_l/λ_e
Virtual object, shown in same optical space as focal figure, enlarged $\lambda_l/\lambda_e \times$
Light source
Pinhole
Condenser lens
γ_m
Hologram
Photographic plate
$z' = z_0\lambda_l/\lambda_e$
$f = L$
Reconstructing lens
OPTICAL SYNTHESIS

FIGURE 1. Principle of electron microscopy by reconstructed wave-fronts.

As the photograph of a diffraction pattern taken in divergent, coherent illumination will be often used in this paper, it will be useful to introduce a special name for it, to distinguish it from the diffraction pattern itself, which will be considered as a complex function. The name 'hologram' is not unjustified, as the photograph contains the total information required for reconstructing the object, which can be two-dimensional or three-dimensional.

The hologram must be either printed, or developed by reversal, and the positive is transferred to the optical synthetizer, which is a light-optical imitation of the electronic device. All essential dimensions, which determine the shape of the wave, are scaled up in the ratio of light wave-length λ_l to electron wave-length λ_e. As electrons of about 50 keV energy, with a de Broglie wave-length of about 0·05 Å, are the most useful in electron microscopy, this ratio will be of the order 100,000. It may be noted that the focal length of the electron lenses is not an essential dimension, and need not be scaled up.

To avoid scaling up the photographic plate a further lens is provided, which enlarges it in the ratio λ_l/λ_e in the optical space of the enlarged focal figure. This means that the image of the hologram is moved practically to infinity, i.e. it must be in the focal plane of the collimator lens. In the illustration it has been assumed for simplicity that the angles are the same in the analyzer and in the synthetizer, but it will be shown later that the condition $f = L$ is not essential. Nor is it necessary to use a separate condenser lens system. The condenser and the collimator, which have been shown separate in figure 1 to simplify the explanations, form one optical unit, whose function it is to produce an imitation of the original wave-front in the plane of the hologram. The spherical aberration, and the practically unavoidable ellipticity of the electron lenses must be reproduced with great accuracy, with a tolerance of about one fringe for the marginal rays.

Thus in the new method it is no longer necessary to correct the spherical aberration of electron lenses. The aperture can be opened up far beyond the limit of tolerance in ordinary electron microscopy. It is only necessary to *imitate* the aberrations to the same accuracy as they would have to be corrected to achieve a certain resolution. Thus the difficulty is shifted from electron optics to light optics, where refracting surfaces can be figured to any shape, without the limitations imposed in electron optics by the laws of the electromagnetic field. On the electron-optical side we require only a certain moderate constancy, sufficient to avoid readjustment of the optical system at too frequent intervals.

The technical difficulties of the scheme will not be dealt with in this paper. It may be only mentioned that they involve mechanical and electrical stability, operation with objects much smaller than those hitherto dealt with in electron microscopy, and the problem of obtaining the high current densities required under the additional condition of coherence. For the rest the paper will deal mainly with the general wave-theoretical foundations of the new method.

THE PRINCIPLE OF WAVE-FRONT RECONSTRUCTION

Consider a coherent monochromatic wave with a complex amplitude U striking a photographic plate. We write $U = A\mathrm{e}^{i\psi}$, where A and ψ are real. U may be decomposed into a 'background wave' $U_0 = A_0\mathrm{e}^{i\psi_0}$, and a remainder $U_1 = A_1\mathrm{e}^{i\psi_1}$ which is due to the disturbance created by the object and may be called the secondary wave. Thus the complex amplitude at the photographic plate is

$$U = U_0+U_1 = A_0\mathrm{e}^{i\psi_0}+A_1\mathrm{e}^{i\psi_1} = \mathrm{e}^{i\psi_0}(A_0+A_1\mathrm{e}^{i(\psi_1-\psi_0)}) \tag{1}$$

and its absolute value $A = [A_0^2+A_1^2+2A_0A_1\cos(\psi_1-\psi_0)]^{\frac{1}{2}}$.

The density of photographic plates, plotted against the logarithm of exposure, is an S-shaped curve, with an approximately straight branch between the two knees. In this region the transmission of intensity is a power $-\Gamma$ of the exposure. The word 'transmission' and the symbol t will be reserved in this paper for the amplitude transmission, which is in general complex; hence the intensity transmission is tt^*, where the asterisk denotes the complex conjugate. For pure absorption, without phase change, t is a real number, the square root of the intensity transmission. Thus we write for the negative process

$$t_n = (K_n A)^{-\Gamma_n},$$

where K_n is proportional to the time of exposure. In the printing of the negative the exposure is proportional to t_n, hence the transmission of the positive print becomes

$$t_p = [K_p(K_n A)^{-\Gamma_n}]^{-\Gamma_p} = K A^{\Gamma}, \tag{2}$$

where $\Gamma = \Gamma_n \Gamma_p$ is the 'overall gamma' of the negative-positive process. The same type of law applies if reversal development is used.

If now in the reconstruction process we illuminate the positive hologram with the background U_0 alone, a 'substituted wave' U_s will be transmitted, which is, apart from a constant factor

$$U_s = U_0 t_p = A_0 \mathrm{e}^{i\psi_0}[A_0^2 + A_1^2 + 2A_0A_1\cos(\psi_1 - \psi_0)]^{\frac{1}{2}\Gamma}. \tag{3}$$

The simplest, and as will be seen also the most advantageous choice, is $\Gamma = 2$, which gives

$$\begin{aligned} U_s = U_0 A^2 &= A_0 \mathrm{e}^{i\psi_0}[A_0^2 + A_1^2 + 2A_0A_1\cos(\psi_1 - \psi_0)] \\ &= A_0^2 \mathrm{e}^{i\psi_0}\left[A_0 + \frac{A_1^2}{A_0} + A_1 \mathrm{e}^{i(\psi_1-\psi_0)} + A_1 \mathrm{e}^{-i(\psi_1-\psi_0)}\right]. \end{aligned} \tag{4}$$

Comparing this with (1) one sees that if $A_0 = \text{const.}$, i.e. if the background is uniform, the substituted wave contains a component proportional to the original wave U (the first and third terms). This is not in itself a proof of the principle of reconstruction, as any wave can be split into a given wave and a rest. It remains to be shown that the remainder, i.e. the spurious part of U_s, does not constitute a serious disturbance.

This remainder consists of two terms. One of these has the same phase as the background, with an amplitude $(A_1/A_0)^2$ times the amplitude of the background. This term can be made very small if the background is relatively strong, which does not mean that the contrast in the hologram must be poor. Assume for instance $(A_1/A_0)^2 = 0{\cdot}01$, i.e. a secondary intensity which is only 1 % of the primary. This gives $A_1/A_0 = 0{\cdot}1$, and the intensity ratio between the maxima and minima of the interference fringes is $(1{\cdot}1/0{\cdot}9)^2 = 1{\cdot}5$. With $\Gamma = 2$ the intensity transmissions will be in the ratio $1{\cdot}5^2 = 2{\cdot}25$, a very strong contrast. The contrast will fall below the observable limit of about 4 % only for $(A_1/A_0)^2 \leqslant 0{\cdot}0001$, i.e. if the flux scattered by the object into the area of the diagram is less than 0·01 % of the illuminating flux. This remarkable effect of the coherent background has been systematically utilized by Zernike (1948) for the amplified display of weak interference fringes.

The second term of the remainder has the same amplitude $A_1 A_0^2$ as the reconstruction of the original secondary wave, but it has a phase shift of opposite sign relative to the background. It may be called for brevity the 'conjugate-complex' wave. The two twin waves carry the same energy.

The conjugate wave produces a serious disturbance only in rather exceptional arrangements; in most cases the twin waves can be effectively separated. To make this plausible one may think of Fresnel-zone plates. These can be, in fact, considered as holograms of a point object, produced by a point source at infinity. Zone plates act simultaneously as positive and negative lenses, producing two focal points, one at each side of the plate, at equal distance, which can be separately observed. As will be shown later, homocentric illumination produces always such twin images, only, with the source at finite distance, these will be in mirror-symmetrical position with respect to the point source, not to the hologram. In beams which are only approximately homocentric the second image is no longer sharp, but effective separation can be always achieved if the object is sufficiently small, and if certain positions are avoided.

While the twin wave cannot be avoided, the spurious term which is proportional to $(A_1/A_0)^2$ and the distortion due to an uneven background can both be eliminated, or at least effectively suppressed by a modification of the photographic process. In the case of small objects, at least over a large part of the hologram, the photographic density difference between two neighbouring interference maxima is insignificant. This makes it possible to wash out the interference fringes by taking a slightly defocused print of the positive hologram, and processing this print with $\Gamma = 1$. If this print, which has a transmission inversely proportional to $A_0^2+A_1^2$, is placed in register with the positive, and illuminated by the background wave U_0, the substituted wave becomes

$$U_s' = A_0 e^{i\psi_0}[A_0^2+A_1^2+2A_0A_1\cos(\psi_1-\psi_0)]/(A_0^2+A_1^2) = e^{i\psi_0}\left[A_0+\frac{2A_1\cos(\psi_1-\psi_0)}{1+(A_1/A_0)^2}\right]$$

$$= e^{i\psi_0}\left[A_0+2A_1\cos(\psi_1-\psi_0)-2\frac{A_1^3}{A_0^2}\cos(\psi_1-\psi_0)+\ldots\right], \qquad (5)$$

in which the spurious term is of the order $(A_1/A_0)^3$ as compared with the background, and the distortion due to a non-uniform background is eliminated. If one only wants to eliminate the background by itself, one can also use a negative photograph taken in the illuminating beam without the object, processed with $\Gamma = 2$.

To discuss briefly also the case $\Gamma \neq 2$, we put for simplicity $A_0 = 1$ and $A_1/A_0 = a$, and obtain from (3) by binomial expansion

$$U_s = e^{i\psi_0}[1+\tfrac{1}{2}\Gamma a^2+\Gamma a\cos(\psi_1-\psi_0)+\tfrac{1}{2}\Gamma(\Gamma-2)\,a^2\cos^2(\psi_1-\psi_0)+\ldots]. \qquad (6)$$

In the reconstructed wave the contrast is enhanced in the ratio $\frac{1}{2}\Gamma$. But, in addition, one obtains twin waves with phase shifts $2(\psi_1-\psi_0)$, etc., but with smaller amplitudes. This makes it evident that $\Gamma = 2$ is the best choice, except if the original contrast is so weak that it must be enhanced, even at the cost of faithfulness in the reproduction.

ILLUMINATION BY A HOMOCENTRIC WAVE

In order to study the reconstruction cycle in more detail, it will be advantageous to start with the simple case of homocentric illumination, which can be approximately realized by a sufficiently small pinhole as light source. It will be convenient to restrict the discussion for a start to two-dimensional objects, occupying a part of some closed surface Σ which encloses the point source O. The object at a point P of Σ may be characterized by an amplitude transmission coefficient $t(P)$, which is the ratio of the complex amplitudes at the two sides of Σ, in proximity of the point P. t is in general a complex datum, real only for purely absorbing objects. It is, of course, understood that the concept of a transmission coefficient, real or complex, is not applicable to an object which is two-dimensional in the mathematical sense. Of a physical object to which this concept is applicable we must assume that it is at least several wave-lengths in thickness. Moreover, we must assume that laterally, in the surface Σ, the function $t(P)$ does not vary appreciably within a wave-length. These are the conditions for the applicability of the Fresnel-Kirchhoff theory of diffraction. In electron optics, operating with fast electrons of about 0·05 Å wave-length, this condition is always satisfied, as there exists no material object (except nuclei) whose physical properties change significantly in less than about ten times this wave-length.

With these qualifications we can apply the Fresnel-Kirchhoff diffraction formula (cf., for example, Baker & Copson 1939, p. 73). The notations are explained in figure 2. If the monochromatic source at O is of unit strength, the amplitude in the illuminating wave is

$$U_0 = \frac{1}{r_0} \mathrm{e}^{ikr_0},$$

where r_0 is the distance measured from O, and $k = 2\pi/\lambda$. The presence of an object in the surface Σ modifies the amplitude at a point Q outside to

$$U(Q) = \frac{1}{2\lambda}\int t(P)\, \mathrm{e}^{ik(r_0+r_1)-\frac{1}{2}\pi i}(\cos\theta_0 - \cos\theta_1)\frac{dS}{r_0 r_1}. \tag{7}$$

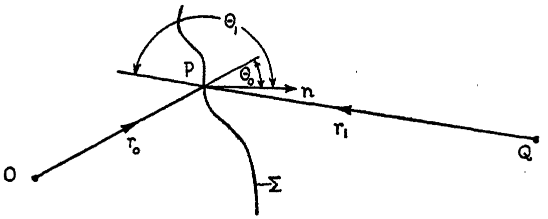


Figure 2. Fresnel-Kirchhoff diffraction formula.

We will now apply this formula to calculate the 'physical shadow' of a plane object at infinity. The 'physical shadow' includes the diffraction effects, and is to be distinguished from the 'geometrical shadow' into which it merges at vanishing wave-length.

As the beams to be used in practice will have semi-cone angles of 0·05 or less, we can put $\cos\theta_0 = -\cos\theta = 1$, and consider the factor $1/r_0 r_1$ as constant. We also drop the constant factor $(1/2\lambda)\,\mathrm{e}^{-\frac{1}{2}\pi i}$, and use equation (7) in the simplified form

$$U(Q) = \int_{\Sigma} t(P)\, \mathrm{e}^{ik(r_0+r_1)}\, dS. \tag{7·1}$$

Using the notations explained in figure 3, the distance r_0 of a point P in the object plane $z = z_0$ is

$$r_0 = (x^2+y^2+z_0^2)^{\frac{1}{2}} = z_0 + \tfrac{1}{2}(x^2+y^2)/z_0 - \tfrac{1}{8}(x^2+y^2)^2/z_0^3 + \dots.$$

In this section we will use only the first two terms of the expansion.

The observation point Q may be at a distance L in the Z-direction, very large compared with z_0, practically at infinity, so that we can write

$$r_1 = L \sec\gamma - (x\cos\alpha + y\cos\beta).$$

The first terms in the expression for r_0 and r_1 give constant phase factors, independent of x, y, which may be dropped. The remaining essential part may be termed 'the amplitude in the direction ξ, η', and is

$$U(\alpha,\beta) = \iint t(x,y)\exp\{ik[(x^2+y^2)/2z_0 - (x\cos\alpha + y\cos\beta)]\}\,dx\,dy. \tag{8}$$

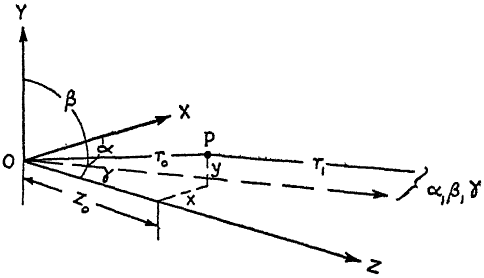


FIGURE 3. Explanation of symbols.

Unless the limits are indicated, integrations in this paper will be always understood to be carried out between infinite limits. As the phase under the integral is valid only for small angles, equation (8) is physically valid only if $t(x, y)$ vanishes rapidly outside a small central area.

It will now be convenient to introduce 'Fourier variables' ξ, η instead of the direction cosines by

$$\xi = \frac{1}{\lambda}\cos\alpha, \quad \eta = \frac{1}{\lambda}\cos\beta. \tag{9}$$

Their connexion with the co-ordinates X, Y in a plane at the large distance $z = L$ is given by

$$X = L\frac{\cos\alpha}{\cos\gamma} = \frac{\lambda L\xi}{[1-\lambda^2(\xi^2+\eta^2)]^{\frac{1}{2}}}, \quad Y = L\frac{\cos\beta}{\cos\gamma} = \frac{\lambda L\eta}{[1-\lambda^2(\xi^2+\eta^2)]^{\frac{1}{2}}}.$$

If the illuminating cone is narrow enough, ξ and η can be taken to represent the co-ordinates in the physical shadow. The geometrical shadow of a point x, y has the Fourier co-ordinates $\xi = x/\lambda z_0$, $\eta = y/\lambda z_0$. The quantity

$$\mu = \lambda z_0 \tag{10}$$

is the only parameter of the diffraction problem. Its square root can be considered as the characteristic length. Details coarser than $\mu^{\frac{1}{2}}$ will be shown to have shadows more or less similar to themselves, finer details lose all likeness by diffraction.

Using the notations (9) and (10) equation (8) can be written, with the abbreviation $x^2+y^2=r^2$,

$$U(\xi,\eta) = \iint [t(x,y)\,e^{\pi i r^2/\mu}]\,e^{-2\pi i(x\xi+y\eta)}\,dx\,dy. \tag{11}$$

Thus the amplitude in the ξ, η direction is the Fourier transform of the function

$$t(x, y)\, \mathrm{e}^{i\pi r^2/\mu}$$

in the standard notation of Campbell & Foster (1931). We can at once write down the reciprocal formula

$$t(x, y) = \mathrm{e}^{-\pi i r^2/\mu} \iint U(\xi, \eta)\, \mathrm{e}^{+2\pi i(x\xi + y\eta)}\, d\xi\, d\eta. \tag{12}$$

It will be useful to study these reciprocal transformations mathematically, while provisionally disregarding the conditions which must be imposed on the function $t(x, y)$ to give them physical validity. First we put them into a more symmetrical form, by imagining the amplitude $U(\xi, \eta)$ as produced by the passage of the illuminating wave U_0 through a 'shadow object' in the plane ξ, η with a transmission $\tau(\xi, \eta)$. (It may be noted that τ is in general complex; thus the shadow object cannot be replaced by a photographic plate.) That is to say, we put

$$U(\xi, \eta) = U_0(\xi, \eta)\, \tau(\xi, \eta). \tag{13}$$

The background U_0 can be obtained directly from (11) by putting $t = 1$

$$U_0(\xi, \eta) = i\mu\, \mathrm{e}^{-\pi i \mu \rho^2}$$

with the abbreviation $\xi^2 + \eta^2 = \rho^2$. We now obtain the symmetrical transformation formulas

$$\tau(\xi, \eta) = \frac{1}{i\mu} \mathrm{e}^{\pi i \mu \rho^2} \iint t(x, y)\, \mathrm{e}^{-\pi r^2/i\mu}\, \mathrm{e}^{-2\pi i(x\xi + y\eta)}\, dx\, dy, \tag{14}$$

$$t(x, y) = i\mu\, \mathrm{e}^{\pi r^2/i\mu} \iint \tau(\xi, \eta)\, \mathrm{e}^{-\pi i \mu \rho^2}\, \mathrm{e}^{+2\pi i(x\xi + y\eta)}\, d\xi\, d\eta. \tag{15}$$

These may be called the 'shadow transformations', and $t(x, y)$, $\tau(\xi, \eta)$ a pair of 'shadow transforms'. They are, of course, intimately related to Fourier transforms, though simpler in some respects.

The transformations (14) and (15) can be derived from one another by the rule: Interchange t and τ, x and ξ, y and η, i.e. interchange Latin and Greek symbols, and replace i by $-i$, μ by $1/\mu$. Two transformations in succession restore the original. Physically this means that if instead of a photograph we could produce a 'shadow object' with the absorbing and refractive properties of $\tau(\xi, \eta)$, and illuminated this with the background, we should exactly restore the object $t(x, y)$ in its original position. As a photograph cannot imitate the imaginary part of τ, a certain residual wave arises, to which we will return in the next section. But it will be useful to consider first a few examples of shadow transforms.

As in the case of Fourier integrals, the transforms of exponentials of quadratic forms are particularly simple and instructive. It is convenient to write these in the form

$$t(x, y) = \exp\left[-\pi(A_1 x^2 + 2B_1 x + A_2 y^2 + 2B_2 y)\right].$$

This is the product of an x- and a y-factor, and as the transform is again the product of a ξ- and an η-factor, it is sufficient to give the transform of

$$t(x) = \mathrm{e}^{-\pi(Ax^2 + 2Bx)}, \tag{16·1}$$

which is $$\tau(\xi) = (1+i\mu A)^{-\frac{1}{2}} \exp\left[-\frac{\pi\mu(\mu A\xi^2+2B\xi-iB^2)}{1+i\mu A}\right]. \tag{16·2}$$

Thus the shadow transform of an exponential of a quadratic form is a function of the same type, as in the case of Fourier transforms, but the relation between the parameters is of a different build. For example, if $A = B = 0$, which makes t a constant, τ will be the same constant, while the Fourier transform of a constant is a delta function, which vanishes everywhere except at the argument zero. Moreover, the shadow transform of a harmonic function ($A = 0$)

$$t(x) = \mathrm{e}^{2\pi ix/p} \tag{16·3}$$

is again a harmonic function $$\tau(\xi) = \mathrm{e}^{-i\pi\mu/p^2}\,\mathrm{e}^{2\pi i\mu\xi/p}. \tag{16·4}$$

The period in the shadow is p/μ, which is *the geometrical shadow of the period* p. The only difference is in the phase factor $\mathrm{e}^{-i\pi\mu/p^2}$. If the period p is long compared with the characteristic length $\mu^{\frac{1}{2}}$, the phase factor tends to unity, which means that if the object contains no details finer than $\mu^{\frac{1}{2}}$ the physical shadow tends towards the geometrical shadow $$\tau(\xi, \eta) \to t(\mu\xi, \mu\eta).$$

Equations (16·3) and (16·4) contain a simple rule for constructing the shadow transform of an object, by expanding $t(x, y)$ into a Fourier integral with periods p_x, p_y. In the transform the Fourier coefficients will differ from the original only in a phase factor $$\exp\left[-i\pi\mu\left(\frac{1}{p_x^2}+\frac{1}{p_y^2}\right)\right].$$

As a practical method this may be used with the cautioning remark that infinite trains of periodic functions are not very suitable for the description of small objects, and that the applicability of equations (14) and (15) to the physical process is strictly speaking limited to objects which transmit appreciably only in a region $x/z_0, y/z_0 \ll 1$.

RECONSTRUCTION WITH HOMOCENTRIC ILLUMINATION

Stigmatic illumination is a particularly simple and instructive illustration of the principle of reconstruction which was broadly explained in the first section. It may be recalled that if the hologram is replaced in the original position and illuminated by the background alone, one obtains in addition to the illuminating or primary wave two other waves, one of which is proportional to the original secondary wave emitted by the object, and the other differs from this only by having phase shifts of opposite sign relative to the background. The other small spurious terms may be disregarded for the moment.

It will now be convenient to subtract the background, i.e. the primary wave, both in the object plane, and in the plane of the photographic plate, and to consider instead of t and τ the functions $$t_1 = t-1 \quad \text{and} \quad \tau_1 = \tau-1. \tag{17}$$

As $t = 1$ corresponds to $\tau = 1$, the functions $t_1(x, y)$ and $\tau_1(\xi, \eta)$ are connected by the relations (14) and (15), the same which connect t and τ. We will talk of t_1 as 'the object proper' and of τ_1 as its shadow.

By equation (6), substituting a photographic plate for the physical shadow means replacing τ_1 by

$$\tfrac{1}{2}\Gamma(\tau_1+\tau_1^*).$$

Substituting this into the inverse shadow transformation (15), we obtain two terms t_1. The first of these differs from the original object proper only in the factor $\tfrac{1}{2}\Gamma$. But in the second term, derived from $\tfrac{1}{2}\Gamma\tau^*$, the sign of i has been reversed, and this results in a spurious figure in the object plane, superimposed on the correct reconstruction of the object.

We can give a simple interpretation to the wave corresponding to τ^* if we observe that in the equation (14) applied to the object proper t_1

$$\tau_1(\xi,\eta) = \frac{1}{i\mu}\mathrm{e}^{i\pi\mu\rho^2}\iint t_1(x,y)\,\mathrm{e}^{-\pi r^2/i\mu}\,\mathrm{e}^{-2\pi i(x\xi+y\eta)}\,dx\,dy, \tag{14·1}$$

reversing the sign of i is equivalent to reversing the sign of x, y and of $\mu = \lambda z_0$ and replacing $t_1(x,y)$ by a function

$$t_1(x,y) = t\ (-x,-y). \tag{18}$$

The transformation has now a parameter $-\mu$ instead of μ, i.e. it corresponds to an object in the plane $-z_0$ instead of in $+z_0$. By equation (18) this object arises from the original by mirroring it on the Z-axis, and changing phase delays into phase advances. Summing up, *the 'twin' wave* τ^* *corresponds to an apparent 'twin object', in central symmetrical position with respect to the point focus* 0, *and with opposite phase-shifting properties.*

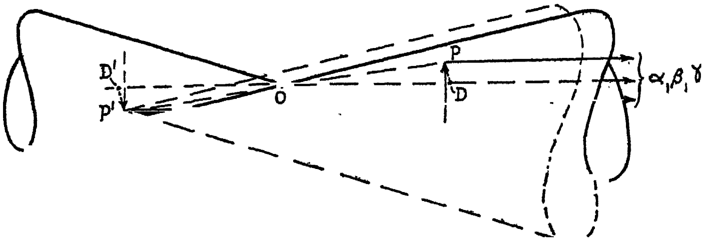


FIGURE 4. The twin images arising in the reconstruction.

Figure 4 is an illustration of the twin objects, from which one can verify this conclusion. The Fresnel-Kirchhoff formula can be interpreted as the sum of elementary spherical waves, originating from the object points P, with amplitudes proportional to $t(P)$. At infinity, in a direction α, β these are plane waves, and their phase difference relative to the background is given by the difference between the ray $\overline{OP}$, and its projection on the direct ray, $\overline{OD}$, apart from the phase shift which arises in the object. Figure 4 makes it clear that the same phase difference, but with opposite sign, would be produced by an object point P', in central symmetrical position to P, if the sign of the phase shift at P' is also reversed.

The interpretation of the residual wave in the reconstruction as a wave emitted by a twin object makes it at once clear that conditions can be found which allow a fairly effective isolation of the reconstructed object, by making use of the limited focal depth of the viewing system. Separation becomes possible if the distance between the twin objects, $2z_0$, exceeds the focal depth D_i, which can be defined as the resolution limit d, divided by the total cone angle utilized in the image formation, $2\gamma_m$. Using Abbe's value d_A for d, the criterion of separation is

$$2z_0 > \frac{d}{2\gamma_m} = \frac{\lambda}{4\gamma_m^2}. \tag{19}$$

If the point focus is not produced by a physical aperture, but by the image of an aperture, formed by an optical system, this is equivalent to the condition that the object must be outside the diffraction region, in which the wave cannot be considered as homocentric.

Outside the focal diffraction region separation is possible, but not complete separation. The twin images will always interfere with one another to some extent, and the interference cannot be regulated at will. This follows from the structure of the transformation equations, which contain only one characteristic length $(\lambda z_0)^{\frac{1}{2}}$, and there is no second length with which to form a dimensionless separation factor. Thus the spurious part of the reconstructed image depends only on the object itself, and on the parameter μ. This disturbance will now be investigated in some detail.

THE SPURIOUS PART OF THE RECONSTRUCTION IN HOMOCENTRIC ILLUMINATION

The simplicity of the transforms (16·1) and (16·2) suggests building up arbitrary plane objects from 'probability spots'. In the limit these tend to two-dimensional delta functions, which can represent any function $t_1(x, y)$, but it is not necessary, nor is it physically justifiable, to pass to this limit. Optical imagery does not operate with points, but with elementary regions of the size of the resolution limit. Inside such a small area the values of $t_r(x, y)$, which describe the reconstituted object, are not independent of one another.

First we carry out the reconstruction cycle for a single probability spot. Assume the transmission in the object plane in the form

$$t(x,y) = 1 - A \exp\left\{-\frac{\pi}{a^2}[(x-x_0)^2+(y-y_0)^2]\right\} = 1 - A\mathrm{e}^{-\pi(r'/a)^2}, \tag{20}$$

where the abbreviation r' has been used for the distance measured from the centre x_0, y_0 of the spot. $(1-A)$ is the amplitude transmitted at the centre of the spot, at unit background. If the object is a pure absorber, A is real, positive, and less than unity. If the object has pure-phase contrast $|1-A| = 1$, and $|A|$ is in the limits 0 to 2.

Equations (16·1) and (16·2) give for the physical shadow of (20)

$$\tau(\xi, \eta) = 1 + \frac{i\epsilon A}{1-i\epsilon}\exp\left(\frac{\pi i\mu\rho'^2}{1-i\epsilon}\right) \tag{21}$$

with the abbreviations $\epsilon = a^2/\mu = a^2/\lambda z_0$

and $\rho'^2 = (\xi - x_0/\mu)^2 + (\eta - y_0/\mu)^2 = (\xi-\xi_0)^2 + (\eta-\eta_0)^2,$

where ξ_0, η_0 is the geometrical shadow of x_0, y_0. The diffraction figure (21) centres around this point. Its character is determined by the dimensionless parameter ϵ. If ϵ is large τ approaches the geometrical shadow $t(\mu\xi, \mu\eta)$. The more important case is $\epsilon \ll 1$, which allows simplifying (21) to

$$\tau(\xi, \eta) \doteq 1 + i\epsilon A\, \mathrm{e}^{-\pi(a\rho')^2 + \pi i \mu \rho'^2}. \tag{21·1}$$

The smaller the original spot, the larger its physical shadow.

The photograph substitutes for the complex physical shadow (21·1) the real transmission function

$$\tau_s(\xi, \eta) = |\tau|^{\Gamma} \doteq 1 + \tfrac{1}{2}\Gamma i\epsilon A\, \mathrm{e}^{-\pi(a\rho')^2}\, \mathrm{e}^{\pi i \mu \rho'^2} - \tfrac{1}{2}\Gamma i\epsilon A^*\, \mathrm{e}^{-\pi i \mu \rho'^2}. \tag{22}$$

This approximation is valid if $\epsilon^2 \ll 1$.

The inverse transformation (15), applied to the first two terms of (22), restores the original object (20), but the contrast is $\frac{1}{2}\Gamma$ times the original. The same transformation applied to the last term of (22) gives the spurious or error term

$$t_e(x, y) = -\tfrac{1}{4}\Gamma i\epsilon A^*\, \mathrm{e}^{-\pi(ar'/2\mu)^2}\, \mathrm{e}^{-\pi i(r'^2/2\mu)}. \tag{23}$$

This is the amplitude (at unit background) which the twin image produces in the original object plane. The spurious image centres on x_0, y_0, but it has a character quite different from the original. The amplitude t_e falls off only slowly with the distance r' from the centre, the slower the smaller the original spot radius a, while the phase changes rapidly, according to the last factor in (23), in a manner independent of the spot size. Thus the spurious image will manifest itself in a system of fine and weak interference fringes, superimposed on the true reconstruction.

The exact value of the reconstructed transmission function t_r in the case $\Gamma = 2$ may be also given for reference:

$$\begin{aligned} t_r(x, y) = 1 - A \exp\left[-\pi\left(\frac{r'}{a}\right)^2\right] - \frac{i\epsilon A^*}{2 + i\epsilon} \exp\left[-\frac{\pi(\epsilon + 2i)}{\mu(4 + \epsilon^2)} r'^2\right] \\ + \frac{\epsilon^2 A A^*}{1 + \epsilon^2 - 2i\epsilon} \exp\left[-\frac{2\pi\epsilon(1 + \epsilon^2 + 2i\epsilon)}{\mu[(1 + \epsilon^2)^2 + 4\epsilon^2]} r'^2\right]. \end{aligned} \tag{22·1}$$

The first two terms stand for the exact reconstruction, the last two for the spurious amplitude. They differ from (23) only in terms of the order ϵ^2 or higher.

The reconstruction cycle in the case $\Gamma = 2$ is illustrated in figure 5 for a probability spot with a black centre. One must be careful not to go beyond $\Gamma = 2$ if there are sharp contrasts in the object. As shown in figure 6, a lighter centre will appear inside a black ring, and black lines will appear doubled.

Up to this point we have assumed unlimited apertures; consequently there was no lower limit to the spot size a which could be correctly reproduced. The effects of limited resolution can be very simply discussed by assuming a mask used in the taking of the hologram, with an amplitude transmission

$$\mathrm{e}^{-\pi(c\rho)^2}.$$

Such graded masks are preferable to sharp apertures, not only from the point of view of mathematical simplicity, but also because they reduce the 'false detail' resulting from the cut-off to a minimum. Their use is well known in structure analysis (Bunn 1945, p. 350).

As the mask is used twice, in the taking of the photograph and in the reconstruction, its total effect is

$$e^{-\pi(\Gamma+1)(c\rho)^2}.$$

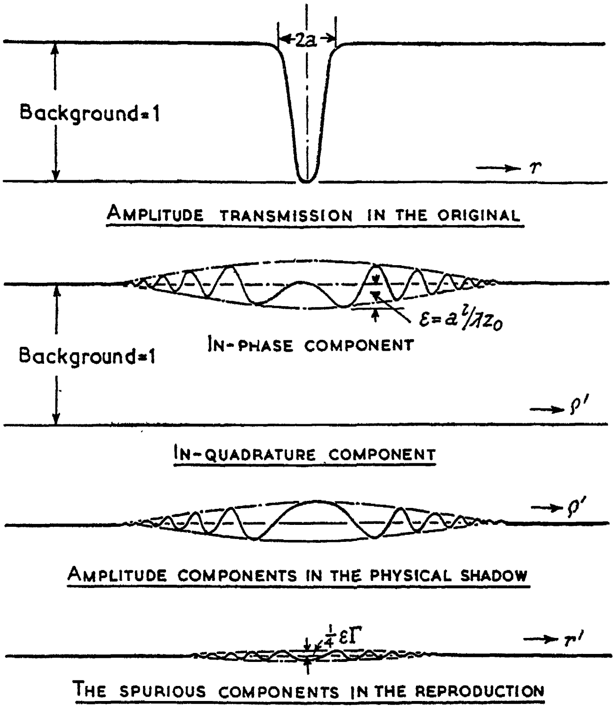

2a
Background=1
r
AMPLITUDE TRANSMISSION IN THE ORIGINAL
ε = a²/λz₀
Background=1
IN-PHASE COMPONENT
ρ′
IN-QUADRATURE COMPONENT
ρ′
AMPLITUDE COMPONENTS IN THE PHYSICAL SHADOW
¼εΓ
r′
THE SPURIOUS COMPONENTS IN THE REPRODUCTION

FIGURE 5. Reproduction cycle of a 'probability spot'.

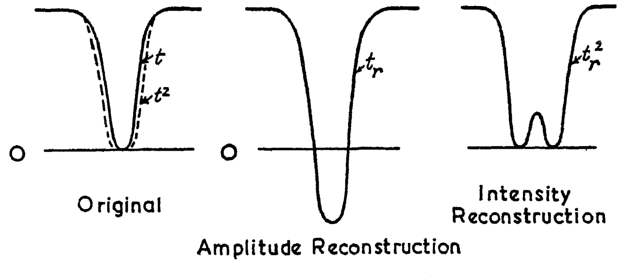


FIGURE 6. Distortion by exaggerated contrast.

We now assume $\Gamma = 2$, so as to obtain correct contrast reconstruction, and we put $3c^2 = b^2$. We have now to reconstruct the object, the probability spot (20), from the physical shadow

$$\tau_s = e^{-\pi(b\rho)^2}\{1 + i\epsilon A\, e^{-\pi(a^2-i\mu)\rho'^2} - i\epsilon A^*\, e^{-\pi(a^2+i\mu)\rho'^2}\}, \tag{24}$$

which differs from (22) only in the masking factor. Introducing the small dimensionless parameter

$$\sigma = b^2/\mu = b^2/\lambda z_0,$$

one obtains by the transformation (15), neglecting powers of ϵ and σ higher than the first,

$$t_r(x,y) = (1+i\sigma)\exp\left[-\frac{\pi\sigma}{\mu}(1+i\sigma)r^2\right] - \frac{A}{1+(b/a)^2}\exp\left\{-\frac{\pi}{a^2+b^2}[r'^2+\epsilon\sigma(x_0^2+y_0^2)]\right\}$$
$$-\tfrac{1}{2}i\epsilon A^*\exp\{-(\tfrac{1}{2}\pi i r'^2/\mu)-(\pi/4\mu)[\epsilon[(x-x_0)^2+(y-y_0)^2]+\sigma[(x+x_0)^2+(y+y_0)^2]]\}. \tag{25}$$

The first is the background term. Apart from a very small diffraction effect, of the order σ^2, it is the geometrical shadow of the mask, projected on the object plane. The second is the 'correct' reproduction term. The chief difference is that the spread of the reproduced spot is now $(a^2+b^2)^{\frac{1}{2}}$ instead of a. Thus b has the meaning of the *resolution limit*, apart from a numerical factor which will be determined later. The factor $[1+(b/a)^2]^{-1}$ before the amplitude expresses the fact that the amplitude decreases in the same ratio as the area of the spot has increased. This loss of contrast for very small objects appears stronger than in the case of ordinary microscopy, where the amplitude falls off with the square root of the area, but the result is the same, as with a strong coherent background the intensity contrast is a linear function of the amplitude.

The error term, in the second row, has a structure different from (23); it no longer centres exactly on the original spot, as it contains a factor which centres on the mirror image of the spot, $-x_0$, $-y_0$. With the abbreviation $r_0^2 = x_0^2+y_0^2$ we can write the error term in a different and very useful form

$$t_e = -\tfrac{1}{2}i\epsilon A^*\exp[-\frac{\pi}{4\mu}(\epsilon-\sigma)r'^2]\exp[-(\tfrac{1}{2}\pi i r'^2/\mu) - (\tfrac{1}{2}\pi\sigma/\mu)(r^2+r_0^2)]. \tag{25·1}$$

This is particularly useful in the case $\epsilon = \sigma$, i.e. $a=b$, as in this case the fringe system t_e has an amplitude independent of r'. The amplitude, though not the phase, centres on $x, y = 0$.

This result can serve as a basis for the theory of the spurious part in the reproduction of arbitrary objects, with the aim of obtaining a criterion for objects suitable for two-step microscopy.

A microscope, like any other optical system, can transmit a finite number of data only. Describing an object by a continuous transmission function is an objectionable idealization, as such a function contains an infinity of non-reproducible detail. We come nearer to an adequate description if we divide up the object by a network of lines into cells of the size of the resolution limit, associate a complex datum with each cell, and investigate the transmission of these data through the optical system.

Equation (25·1) suggests that particularly simple results will be obtained if we represent the object by a two-dimensional lattice of probability spots with a spread $a = b$. As illustrated in figure 7, we arrange these spots in a hexagonal lattice, with a distance d between adjacent centres. d is the resolution limit, which we define in a way slightly different from the usual, by postulating that three (instead of two) equal probability spots with spreads $a = b$ can just be resolved if their centres are at distances d from one another, i.e. the minimum in the centre just vanishes. By equation (25) the amplitude in the correct reproduction term follows a law $e^{-\frac{1}{2}\pi(r'/b)^2}$ if $a = b$. In the middle between three centres $r' = d/\sqrt{3}$; thus the condition for d is

$$\exp\left[-\frac{\pi}{2}\left(\frac{d}{\sqrt{3}b}\right)^2\right] = \frac{1}{3},$$

which gives
$$d = 1{\cdot}45b.$$

This is in good agreement with the usual definition of the resolution limit,

$$d = 0{\cdot}6\lambda/\sin\gamma_m,$$

if we define γ_m as the angle at which the background amplitude has dropped to $1/\sqrt{3}$, i.e. the background intensity to $1/3$ of its maximum value. Denoting the corresponding radius in the object plane by $R = z\sin\gamma_m$, we have

$$\exp\left[-\frac{2\pi\sigma}{\mu}R^2\right] = \exp\left[-2\pi\left(\frac{b}{\lambda}\sin\gamma_m\right)^2\right] = \frac{1}{3},$$

which gives
$$b = 0{\cdot}42\lambda/\sin\gamma_m, \quad d = 0{\cdot}61\lambda/\sin\gamma_m.$$

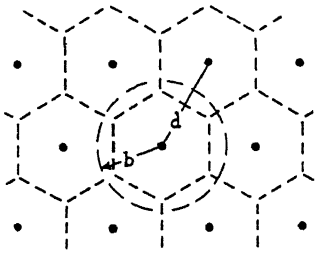


FIGURE 7. Dividing the object into independent elements.

Call N the number of independent elements inside the illuminated field, i.e. number of cells contained in the disk of radius R. Each cell occupies an area

$$\tfrac{1}{4}\sqrt{3}d^2 = 0{\cdot}433d^2 = 0{\cdot}91b^2;$$

thus the number N is

$$N = \frac{\pi}{0{\cdot}91}\frac{R^2}{b} = 3{\cdot}45(0{\cdot}42)^2\left(\frac{\lambda z_0}{b^2}\right)^2 = 0{\cdot}61\frac{1}{\sigma^2}. \tag{26}$$

N can be easily made a very large number, of the order 10^6 to 10^8. This suggests a *statistical evaluation* of the spurious part of the reproduction, by assuming random distribution of the amplitude over the independent elements of the object. It is, of course, understood that this might lead to gross errors in special cases, but it is certainly an acceptable assumption if a great number and variety of objects are considered.

Number the elements from 1 to N. The spurious amplitude in the reconstruction at a point x, y results from the superposition of the error terms of the form (25·1), one for each cell with centres x_n, y_n. The distance of x, y measured from x_n, y_n may be called r'_n. With the simplification resulting from $\sigma = \epsilon$ the resulting error amplitude is

$$t_e(x,y) = \tfrac{1}{2}i\sigma\, e^{-\pi\sigma r^2/2\mu}\sum^{N} A_n^*\, e^{-\frac{1}{2}\pi\sigma r_n^2/\mu}\, e^{-\frac{1}{2}\pi i r_n'^2/\mu}, \tag{27}$$

where we have written r_n for the distance of x_n, y_n from $0, 0$.

The relative distance r'_n between x, y and x_n, y_n occurs here only in the phase factor $\mathrm{e}^{-\frac{1}{2}\pi i r'_n/\mu}$. The two probability decay factors fall off slowly. The first of these, before the sum, is the square root of the background attenuation, i.e. it falls off at half the rate of the background amplitude. The second factor centres on 0, 0 and falls off at the same rate as the first factor. Thus it is admissible to put both equal to unity as an approximation, and simplify equation (27) to

$$t_e(x,y) = \tfrac{1}{2}i\sigma \sum^{N} A_n^* \,\mathrm{e}^{-\frac{1}{2}\pi i r_n'^2/\sigma}, \tag{27·1}$$

that is to say, in order to obtain the error amplitude at x, y we have to superimpose at this point a large number N of *undamped* waves, with wave-lengths $2\mu/r'$, emanating from all image points x_n, y_n. This wave-length is always longer than the resolution limit d. Its smallest value is at $r' = 2R$ and is μ/R, while the resolution limit is $0{\cdot}61\mu/R$.

Introduce now the hypothesis that there is no correlation between the phases of these waves. With this assumption the mean square of the component of t_e in phase with the background, which may be called $t^2_{\text{eff.}}$, is one-half of the sum of the absolute squares of the terms at the right:

$$t^2_{\text{eff.}} = \tfrac{1}{8}\sigma^2 \sum^{N} A_n A_n^* = \tfrac{1}{8}\sigma^2 N \overline{A_n A_n^*} = \tfrac{1}{8}\sigma^2 N A^2_{\text{eff.}}. \tag{28}$$

Here we have introduced the notation $A^2_{\text{eff.}}$ for the mean square secondary amplitudes, $\overline{A_n A_n^*}$, averaged over the whole field. It is understood that the average level of transmission of the object has to be considered as part of the background, and $A_{\text{eff.}}$ is a measure of the departure from uniformity. Combining (28) with equation (26), $N\sigma^2 = 0{\cdot}61$, we obtain

$$t_{\text{eff.}} = 0{\cdot}28 A_{\text{eff.}}. \tag{29}$$

Equation (29) enables us to formulate a criterion for suitable objects. A background can be considered as practically even if the intensity contrast does not exceed about 5 %, i.e. if the amplitude contrast is less than 2·5 %. This means that for suitable objects we must have

$$A_{\text{eff.}} \leqslant 0{\cdot}1 \tag{30}$$

averaged over the whole field. As an example consider a black-and-white object in which the black part, where $A = 1$, covers a fraction κ of the illuminated field, while for the rest $A = 0$. In this case $A_{\text{eff.}} = \sqrt{\kappa}$, and we obtain the simple rule that not more than about 1 % of the illuminated field should be covered with black dots or lines. If, for instance, the object is a disk, half black and half white, its diameter should not exceed one-seventh of the field diameter.

As a second example consider an object with pure phase contrast, but with random distribution of phase delays. We must qualify this by the condition which precedes every application of the Fresnel-Kirchhoff theory; the phase must not vary appreciably between points spaced at less than a wave-length. In other words, the object must appear even and transparent if it is sharply focused. In ordinary microscopy a crinkled sheet of celluloid, or even reticulated gelatine, will satisfy this condition, but not an opal glass with colloidal dispersion. With this qualification in mind we can apply equation (29) and it can be shown that the value of $A_{\text{eff.}}$ is again unity. In the case of pure phase contrast the complex transmission vector $t = 1 - A$ moves on the unit circle, all orientations of t are equally probable. Hence $\bar{t} = 0$, which makes $\bar{A} = 1$, and

$$A^2_{\text{eff.}} = \overline{|A-\bar{A}|^2} = \overline{|A-1|^2} = |t|^2 = 1.$$

This means that an object of this type, if it covers the whole field, produces $t_{\text{eff.}} = 0{\cdot}28$, a very serious disturbance. This result is of interest, because it shows that an irregular transparent support for the object, even if it would be invisible in ordinary microscopy, will make all but the most contrasty or regular features of the object invisible. As it is rather doubtful whether an 'optically flat' or at least acceptably regular supporting membrane can be found in electron microscopy, it appears preferable to use supporting membranes only in a small fraction of the field, or to dispense with them altogether.

IMPROVING THE SEPARATION BY MASKING AND OTHER MEANS

These results lead to the conclusion that high-grade purity in the reproduction cannot easily be achieved even with very small objects, as the spurious intensity is proportional to the square root of the object area. But in the case of small objects special techniques become available, which allow a very effective elimination of the spurious amplitudes. The first of these is the masking of the geometrical shadow in the hologram. The second technique is the masking of the background in the reconstruction process.

The spurious amplitude is objectionable only in the area occupied by the true image. Thus we need eliminate only those rays issuing from the twin object which pass through the object area. As may be seen from figure 4, if the object is small these rays will have substantially the same direction as the primary rays which illuminate the object. This means that we can substantially reduce the spurious amplitude if we mask out the geometrical shadow in the hologram.

This masking process, however, will introduce two new disturbances. First, the mask itself will produce a system of interference fringes. This effect can be reduced to a very low level if a 'probability mask' is used. Secondly, the mask will eliminate some of the data required for a complete reconstruction. Evidently the coarser detail will suffer most, as this is contained in or near to the geometrical shadow area in the hologram, while the finer detail is spread over a larger area outside. But if the object is of the order of the characteristic length $\mu^{\frac{1}{2}}$ or smaller, the suppressed detail becomes insignificant. Thus masking of the shadow is a very effective method for improving the reproduction of very small objects.

In the second method the background, i.e. the primary wave, is suppressed *after* it has traversed the hologram. This can be done by producing a real image of the point source by means of the reconstructing lens in figure 1, and arranging a small black mask at this point, preferably a probability mask. This arrangement is similar to the well-known 'schlieren' method. The result is, that instead of an amplitude in the object plane

$$1 - t_c - t_e,$$

where c stands for 'correct' and e for 'error', we now obtain

$$-t_c - t_e,$$

neglecting the diffraction effects at the mask. Hence an absorbing object will now appear bright on a dark background, as in 'dark-field illumination'. While in the ordinary or 'bright field' method the intensity is approximately

$$1-2t_c-2t_e,$$

the 'dark field' intensity is $\quad t_c^2+2t_et_c+t_e^2.$

One can consider t_e^2 as the spurious background, while $2t_et_c$ is the interference product of the two images. The spurious background is now the square of its previous value, proportional to the coverage fraction instead of to its square root, and becomes negligible for objects which cover only a few percent of the illuminated field. There remains, however, the interference product $2t_et_c$. This contributes nothing to the background, as it is zero everywhere outside the object, where $t_c=0$. In the object area it represents merely a small modulation of the correct density values. In the case of black-and-white objects this effect is negligible, as the outlines remain unchanged. How far it can distort graded objects is a matter for further investigation.

A combination of the two methods, i.e. masking the geometrical shadow *and* the primary wave, appears to be particularly promising in the case of small objects.

A third, somewhat laborious method for improving the separation is taking a series of reconstructions, with different values of μ. While the true image always remains the same, the spurious image varies, and can thus be discriminated. A fourth method will be discussed later, in connexion with non-homocentric illumination.

ILLUMINATING WAVES WITH ASTIGMATISM AND SPHERICAL ABERRATION

Following a method first introduced by Debye, we build up a general coherent illuminating wave of plane wavelets, normal to the direction α,β,γ, with an amplitude $Ad\Omega$ in the infinitesimal solid angle $d\Omega$:

$$A(\alpha,\beta)\exp\{ik[x\cos\alpha+y\cos\beta+z\cos\gamma-p(\alpha,\beta)]\}\,d\Omega. \tag{31}$$

The amplitude A is assumed as real, the phase factor e^{-ikp} expressing the advance of the phase compared with the direct ray through the origin O. Assuming that O coincides with the 'mean paraxial focus' of the beam, let the phase function p be

$$p(\alpha,\beta)=\tfrac{1}{2}A_s(\cos^2\alpha-\cos^2\beta)+\tfrac{1}{4}(C_x\cos^4\alpha+2C_{xy}\cos^2\alpha\cos^2\beta+C_y\cos^4\beta). \tag{32}$$

The first term is the phase advance due to astigmatism, the second is 'elliptical' spherical aberration. It has been assumed for simplicity that the elliptical errors of second and fourth order have the same principal axes x, y.

With the polar angles γ, θ, connected with α,β by

$$\cos\alpha=\sin\gamma\cos\theta,\quad \cos\beta=\sin\gamma\sin\theta, \tag{33}$$

p can be put into the form

$$\begin{aligned}p(\gamma,\theta)&=\tfrac{1}{2}A_s\sin^2\gamma\cos 2\theta\\&\quad+\tfrac{1}{4}\sin^4\gamma[\tfrac{3}{8}(C_x+C_y)+\tfrac{1}{4}C_{xy}+\tfrac{1}{2}(C_x-C_y)\cos 2\theta-\tfrac{1}{4}[C_{xy}-\tfrac{1}{2}(C_x+C_y)]\cos 4\theta].\end{aligned} \tag{34}$$

The fourth-order term now appears as the sum of spherical aberration and two astigmatism terms, one elliptical, the other with fourfold periodicity. If the lens is round

$$C_x = C_y = C_{xy} = C_s, \tag{35}$$

and the fourth-order astigmatic terms vanish. C_s is the constant of spherical aberration. Its meaning is illustrated in figure 8, which shows the ray structure of a beam. The geometric-optical approximation is well justified in the most important practical applications of the present theory, as it is proposed to use beams with apertures about ten times larger than in ordinary electron microscopy, where the diffraction disk is of the same order as the geometrical aberrations. As the minimum cross-section of the beam increases with the third power of the aperture, and the diffraction effect is inversely proportional to the first power, it will represent a small correction only, of the order of 10^{-4} of the geometrical dimensions.

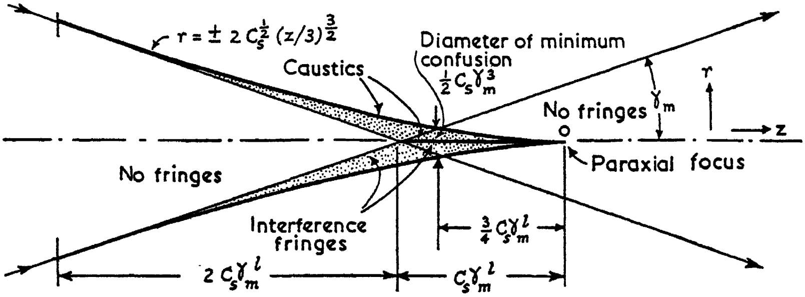

$r = \pm 2C_s^{\frac{1}{2}}(z/3)^{\frac{3}{2}}$
Caustics
Diameter of minimum confusion
$\frac{1}{2}C_s\gamma_m^3$
No fringes
γ_m
r
z
O
Paraxial focus
No fringes
Interference fringes
$\frac{3}{4}C_s\gamma_m^l$
$2C_s\gamma_m^l$
$C_s\gamma_m^l$

FIGURE 8. Focal figure of a beam with spherical aberration C_s.

If the aperture angle is γ_m, all rays cross the axis in the axial caustic, a line of length $C_s\gamma_m^2$ behind the paraxial focus O. The diameter of the beam in the Gaussian plane $z = 0$ is $2C_s\gamma_m^3$, but at the cross-section of minimum confusion, at $z = -\frac{3}{4}C_s\gamma_m^2$, it is four times less. The minimum cross-section is the intersection of the envelope or outer caustic, a rotational surface with an equation $r = \pm 2C_s^{\frac{1}{2}}(z/3)^{\frac{3}{2}}$, with the cone of maximum divergence, $r = \pm(z + C_s\gamma_m^2)\gamma_m$. This cone and the outer caustic divide up the beam into four regions of different character, of which two are dotted, to indicate that they contain interference fringes. The first of these is inside the envelope but outside the cone. The rays cross in every point of this region. The second is the region surrounding the axial caustic, limited by the envelope and by the cone of maximum divergence, which has three rays crossing in every point. The interference fringes in both regions are so sharp and contrasty as to make objects placed into them almost invisible; thus the whole dotted volume must be ruled out as a possible location for objects. In the remaining two regions, at the right and left, there is only one ray to every point, and they represent smoothly graded backgrounds, suitable for microscopic objects. In the region at the left the illumination density is largest near the edge; in the second, at the right, the density has a maximum on the axis.

If, in addition, as is always the case in electron optics, the beam is also astigmatic, figure 8 can still serve as an illustration, but only for the principal sections of the beam, and these must be imagined as displaced longitudinally by $\pm A_s$. Thus O will be now in the middle between the two focal lines, at right angles to one another and to the beam axis, and a distance $2A_s$ apart.

Returning to the wave-optical representation, summing the wavelets (31) gives for the complex amplitude at the point x, y, z

$$U_0(x,y,z) = \iint A(\gamma,\theta)\exp\{ik[(x\cos\theta + y\sin\theta)\sin\gamma + z\cos\gamma - \tfrac{1}{2}A_s\sin^2\gamma\cos 2\theta - \tfrac{1}{4}C_s\sin^4\gamma]\}\sin\gamma\, d\gamma\, d\theta. \quad (36)$$

We have used here the simplifications arising from equation (35), and these will be assumed also in the following formulae to simplify the discussion, but the results will be of such a nature as to permit their extension without difficulty also to the more general case expressed by equations (32) and (34).

Introduce under the integral sign in (36) the Fourier variables $\xi = (\cos\alpha)/\lambda$, $\eta = (\cos\beta)/\lambda$ and $\rho = (\xi^2+\eta^2)^{\frac{1}{2}}$. The exact transformation equations are

$$\sin\gamma = \lambda\rho, \quad \cos\gamma = (1-\lambda^2\rho^2)^{\frac{1}{2}}, \quad d\Omega = \sin\gamma\, d\gamma\, d\theta = \frac{\lambda^2\, d\xi\, d\eta}{(1-\lambda^2\rho^2)^{\frac{1}{2}}}.$$

We again assume a sufficiently narrow beam to justify neglecting $\sin^2\gamma = \lambda^2\rho^2$ in the denominator of the last expression. In the phase factor, however, we must take into consideration terms up to the fourth order in ρ, and write

$$\cos\gamma = (1-\lambda^2\rho^2)^{\frac{1}{2}} = 1 - \tfrac{1}{2}\lambda^2\rho^2 - \tfrac{1}{8}\lambda^4\rho^4.$$

In this approximation

$$U_0(x,y,z) = \lambda^2 e^{ikz}\iint A(\xi,\eta)\exp\{2\pi i[x\xi + y\eta - \tfrac{1}{2}z\lambda\rho^2 - \tfrac{1}{2}A_s\lambda(\xi^2-\eta^2) - \tfrac{1}{8}(z+2C_s)\lambda^3\rho^4]\}\, d\xi\, d\eta. \quad (36{\cdot}1)$$

This integral, like the exact expression (36), can be easily evaluated at large distances R from the origin, in a direction α, β. One obtains

$$U_0(R,\alpha,\beta) = -i\frac{\lambda}{R}A(\gamma,\theta)\, e^{ik(R-p)}, \quad (36{\cdot}2)$$

where p is given by equations (32) or (34). The factor $-i$, which expresses an advance of the wave-front by $\frac{1}{4}\lambda$ as compared with the components (31) arises in the transition from plane to spherical waves, and is familiar in diffraction theory. Equation (36·2) supplies the background to the physical shadow of an object, which we are now going to calculate.

The object, in a plane $z = z_0$, may be characterized, as before, by the complex transmission function $t(x,y)$. Using the fundamental premissa of the Fresnel-Kirchhoff diffraction theory, we assume that the amplitude immediately before the object is that of the undisturbed illuminating wave, $U_0(x,y,z_0)$, and $t(x,y)$ times this immediately behind it. We must now give the angular variables in the illuminating wave by suffixes '0' ('original') to distinguish them from the variables of the outgoing wave, without suffixes.

The problem is building up the outgoing wave from the diffraction products of the plane wavelets which compose the original wave. The Fresnel-Kirchhoff formula in the simplified form (7·1) can be again applied, but with the modification that the wave $r_0^{-1}\,\mathrm{e}^{ikr_0}$ must now be replaced by the sum of the wavelets (34). There is no change in the meaning of r_1, the distance of the observation point $Q(R,\alpha,\beta)$ from the object point P. Thus the Fresnel-Kirchhoff formula now assumes the form

$$U(Q) = U(R,\alpha,\beta)$$

$$= \frac{1}{i\lambda R}\iiiint t(x,y)\,A(\alpha_0,\beta_0)\exp\{ik[r_0(\alpha_0,\beta_0)+r_1(\alpha,\beta)]\}\,dx\,dy\,d\alpha_0\,d\beta_0, \quad (37)$$

where
$$r_0(\alpha_0,\beta_0) = x\cos\alpha_0 + y\cos\beta_0 + z_0\cos\gamma_0 - p(\alpha_0,\gamma_0),$$
$$r_1(\alpha,\beta) = R - x\cos\alpha - y\cos\beta - z_0\cos\gamma.$$

Expressing the angles by the Fourier variables ξ,η and ξ_0,η_0 we obtain, with the same approximations as in (36·1),

$$U(R,\xi,\eta) = \frac{\lambda}{iR}\,\mathrm{e}^{ikR}\iint t(x,y)\,dx\,dy$$
$$\iint A(\xi_0,\eta_0)\exp\{2\pi i\,[x(\xi_0-\xi)+y(\eta_0-\eta)-\tfrac{1}{2}z_0\lambda(\rho_0^2-\rho^2)-\tfrac{1}{8}z_0\lambda^3(\rho_0^4-\rho^4)$$
$$-\tfrac{1}{2}A_s\lambda(\xi_0^2-\eta_0^2)-\tfrac{1}{4}C_s\lambda^3\rho_0^4]\}\,d\xi_0\,d\eta_0. \quad (38)$$

The symmetry of this expression is disturbed by the last two terms, but it is at once restored if we go over to the 'physical shadow', by dividing the amplitude $U(R,\alpha,\beta)$ into the background $U_0(R,\alpha,\beta)$ as given by equation (36·2):

$$\tau(\xi,\eta) = \iiiint t(x,y)\,\frac{A(\xi_0,\eta_0)}{A(\xi,\eta)}\exp[2\pi i\{x(\xi_0-\xi)+y(\eta_0-\eta)-\tfrac{1}{2}z_0\lambda(\rho_0^2-\rho^2)$$
$$-\tfrac{1}{2}A_s\lambda[(\xi_0^2-\xi^2)-(\eta_0^2-\eta^2)]-\tfrac{1}{8}\lambda^3(z_0+2C_s)(\rho_0^4-\rho^4)\}]\,dx\,dy\,d\xi_0\,d\eta_0. \quad (39)$$

This is the formula for the physical shadow at infinity of an object at $z = z_0$, illuminated by a beam with fourth-order aberrations, but which can be evidently extended to aberrations of any order. It is the equivalent of the transformation formula (14) for homocentric illumination, but it cannot be put into the form of an integral over the object plane, as the integration over the angular variables cannot be carried out in terms of the transcendentals recognized in analysis. On the other hand, it can be immediately reduced to a double integral over the angular variables by means of the Fourier transform $T(\xi,\eta)$ of $t(x,y)$ which is

$$T(\xi,\eta) = \iint t(x,y)\,\mathrm{e}^{-2\pi i(x\xi+y\eta)}\,dx\,dy,$$

which converts equation (42) into

$$\tau(\xi,\eta) = \iint T(\xi-\xi_0,\eta-\eta_0)\,\frac{A(\xi_0,\eta_0)}{A(\xi,\eta)}\exp[\pi i\{z_0\lambda(\rho^2-\rho_0^2)$$
$$+A_s\lambda[(\xi^2-\xi_0^2)-(\eta^2-\eta_0^2)]+\tfrac{1}{4}\lambda^3(z_0+2C_s)(\rho^4-\rho_0^4)\}]\,d\xi_0\,d\eta_0. \quad (40)$$

This transformation may be illustrated by a few simple examples. If $t = 1$, i.e. if there is no object, T is a delta function

$$T(\xi-\xi_0,\eta-\eta_0) = \delta(\xi-\xi_0,\eta-\eta_0),$$

which means that the integral (40) is the value of the integrand for $\xi_0 = \xi$, $\eta_0 = \eta$, which is unity, as before.

If $t(x, y)$ is a harmonic function of x, y with periods $1/a, 1/b$

$$t(x, y) = e^{2\pi i(ax+by)}. \tag{41·1}$$

T is again a delta function, but shifted to the point a, b

$$T(\xi-\xi_0, \eta-\eta_0) = \delta(\xi-\xi_0-a, \eta-\eta_0-b),$$

and we have again to take the value of the integrand, but this time at $\xi_0 = \xi-a$, $\eta_0 = \eta-b$. The physical shadow is

$$\tau(\xi, \eta) = \frac{A(\xi-a, \eta-b)}{A(\xi, \eta)} \exp\{\pi i a(2\xi-a)[\lambda(z_0+A_s)+\tfrac{1}{4}\lambda^3(z_0+2C_s)(2\xi^2-2\xi a+a^2)]\}$$
$$\exp\{\pi i b(2\eta-b)[\lambda(z_0-A_s)+\tfrac{1}{4}\lambda^3(z_0+2C_s)(2\eta^2-2\eta b+b^2)]\}. \tag{41·2}$$

We have met the first factors under the exponential in the shadow transformation with homocentric illumination. But the period in the shadow is no longer a constant, that is to say, the shadow of a sinusoidal grid is not of the same type as the original. If, for example, $b = 0$, i.e. the grid is parallel to y, the spacing between two maxima is

$$1/(\lambda z_0 a)[1+A_s/z_0+\tfrac{1}{4}\lambda^2(1+2C_s/z_0)(2\xi^2-2\xi a+a^2)].$$

The first factor is the geometrical shadow of the period $1/a$, the second is the correction arising from astigmatism and spherical aberration, and also from the fourth-order term which expresses the departure of a spherical wave-front from a paraboloid. In all practical applications z_0 will be of the order $C_s\gamma_m^2$, and z_0 can be neglected against $2C_s$. Thus the astigmatism and spherical aberration of a beam can be determined from two holograms of a sinusoidal grid, taken in two positions, at right angles to one another. But the method is not very sensitive. Near the edge of the field where $\xi \gg a, \eta \gg b$, the spacing of two neighbouring maxima will be a fraction

$$1\Big/\Big(1+\frac{C_s}{z_0}\gamma_m^2$$

of the geometrical spacing. But as z_0 will be of the order $C_s\gamma_m^2$ if good photographs are to be obtained, this fraction will be of the order unity. This shows that a sinusoidal grid, even if it were available, would not be a very suitable test object. Spherical aberration can be much better determined from the physical shadow of a thin wire, but the discussion of this case cannot be carried out in elementary terms, and may be omitted.

RECONSTRUCTION IN THE PRESENCE OF SPHERICAL ABERRATION AND ASTIGMATISM

Assume that a photograph has been taken of the physical shadow of an object, according to equations (39) or (40). We have seen that, if the background is relatively strong, this is equivalent to substituting for τ_1 its real part, $\frac{1}{2}(\tau_1+\tau_1^*)$, where, as before, τ_1 relates to the 'object proper' without the background. In order to find the spurious term in the reconstructed object, we must apply to τ_1^* the transformation inverse to (39). But this is rather complicated, while an interpretation in terms of 'twin images' is easy, and leads to much simpler and clearer results.

An expression for τ_1^*, the complex conjugate of the physical shadow τ_1 is obtained from (39) by reversing the sign of i. Assume now, as before, in the plane $z = z_0$ a twin object with a transmission function

$$t_1'(x, y) = t_1^*(-x, -y).$$

Renaming the integration variables $-x, -y$ instead of x, y, one obtains for τ_1^* the expression

$$\tau_1^*(\xi, \eta) = \iiiint t_1'(x, y) \frac{A(\xi_0, \eta_0)}{A(\xi, \eta)} \exp\{2\pi i[x(\xi_0 - \xi) + y(\eta_0 - \eta) + \tfrac{1}{2} z_0 \lambda (\rho_0^2 - \rho^2) + \tfrac{1}{2} A_s \lambda [(\xi_0^2 - \xi^2) - (\eta_0^2 - \eta^2)] + \tfrac{1}{8} \lambda^3 (z_0 + 2C_s)(\rho_0^4 - \rho^4)]\} \, dx\, dy\, d\xi_0\, d\eta_0. \quad (42)$$

This is the physical shadow of an object t_1' in the plane $-z_0$, according to equation (39), but with the important difference that the sign of A_s and C_s has been also reversed. The physical significance of this becomes clearer if instead of τ_1^* we consider the complementary wave U_1' which arises in the reconstruction, and which is obtained from (42) by multiplying it with the background (36·2). The result can be written

$$U_1'(R, \xi, \eta) = \frac{\lambda}{iR} e^{ikR} \iint t_1'(x, y)\, dx\, dy \iint A(\xi_0, \eta_0) \exp\{2\pi i [x(\xi_0 - \xi) + y(\eta_0 - \eta) + \tfrac{1}{2} z_0 \lambda (\rho_0^2 - \rho^2) + \tfrac{1}{8} z_0 \lambda^3 (\rho_0^4 - \rho^4) + \tfrac{1}{2} A_s \lambda (\xi_0^2 - \eta_0^2) + \tfrac{1}{4} C_s \lambda^3 \rho_0^4]\} \times \exp\{-2\pi i [A_s \lambda (\xi^2 - \eta^2) + \tfrac{1}{2} C_s \lambda^3 \rho^4]\}\, d\xi_0\, d\eta_0. \quad (43)$$

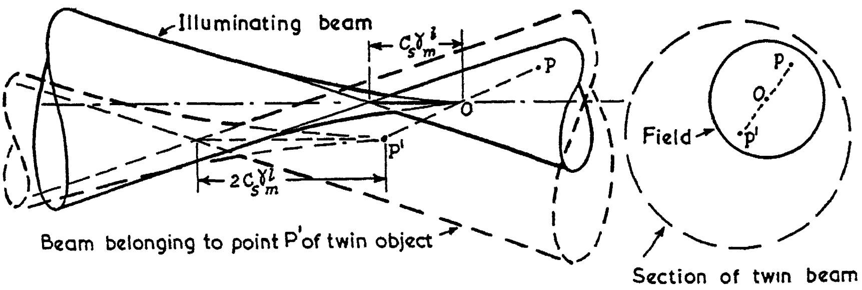


FIGURE 9. The twin object in a beam with spherical aberration.

Comparing this with equation (41), it can be seen that the first two lines represent the emission of an object t_1' in the plane $-z_0$, but *illuminated by a wave in which the signs of the astigmatism A_s and of the spherical aberration C_s are reversed.* This assures complete symmetry in the illumination of the object and its twin. But the emitted wave is modified by the phase factor in the last line. This means that the wavelet issuing from any element $t_1'(x, y)\, dx\, dy$ of the twin object has astigmatism $2A_s$ and spherical aberration $2C_s$. Thus *in the presence of astigmatism or spherical aberration the twin object which appears in the reconstruction will be no longer sharp, but will appear as if viewed through a system with twice the aberrations of the condenser system.* One could, of course, view the twin object instead of the original by means of a viewing system with aberrations of the opposite sign, but not both simultaneously.

This result is illustrated in figure 9, which allows also an elementary verification. The illuminating beam envelope is shown in continuous lines, the beam appearing to issue from a point P' of the twin object in interrupted lines. The axial caustic of this beam is always twice the caustic of the illuminating beam. This can be immediately understood if one imagines the axial caustic as the locus of the centres of homocentric beams, each emitting rays only in a certain cone. For each of these partial beams there exists a sharp twin point to P, on the line joining P with its centre. Equation (43) proves that this geometric-optical reasoning is in fact justified.

Figure 9 shows also that the beam associated with any point of the twin object intersects the object plane in an area four times larger than the field. From this we can infer at once that if the illumination were even, the spurious amplitude in the object plane would bear the same relation to the correct amplitude as in the case of homocentric illumination, i.e. equation (29) would apply again. In fact the illumination is very uneven in a beam with spherical aberration in cross-sections not very far from the caustic, and on this is based a fourth method of improving the separation, in addition to the others which have been discussed in a previous section. Masking is not very efficient in the presence of spherical aberration, as the geometrical shadow of a point object is a radial line, the projection of the axial caustic. This becomes small only if the object is in the axis, but in electron optics it is not possible to fix small objects by means of a transparent support in the middle of the field.

This fourth method for improving the separation is to place the object in a position where it receives less than the average of illumination density. To explain this briefly, define as 'coefficient of illumination', J, the ratio of the mean intensity over a small object area to the mean over the *whole* illuminated field. If the object has the average intensity transmission tt^* and covers a fraction κ of the field, the fraction of the total flux issuing from the object is $tt^*\kappa J$. Exactly the same flux emerges also from the twin object. But of this only a fraction $\frac{1}{4}\kappa J$ will pass through the object. The factor J is here the same as defined from the direct illumination of the object, because, as may be seen in figure 9, the small twin objects interfere with one another in the direction in which they are directly illuminated. Passing from the intensities to the root mean square amplitudes one obtains a separation factor proportional to $\sqrt{(\kappa J)}$, i.e. $\sqrt{J}$ times what we have previously obtained for uniform, homocentric illumination. Thus by placing small objects in *relatively dark* parts of the field, where $J < 1$, one improves the separation, by reducing the spurious background in the object area. Correspondingly more light is sent by the twin object to other regions of the field, but the spurious amplitude is of course harmless if it falls well outside the reconstructed object.

It may be noted that relatively weak illumination does not affect the contrast in the reconstructed object, so long as it is not submerged by ghosts, scattered light, and impurities arising from uneven development of the photograph.

COHERENCE CRITERIA

Up to this point we have assumed an absolutely coherent monochromatic illuminating wave, originating from a point source, but distorted by passing through a lens system. Absolute coherence means interference fringes of any order, but it means of course zero intensity. In practice we must strike a compromise between these two conflicting claims. The best compromise is obtained if the degree of coherence is just sufficient to produce an interference pattern from which the object can be reconstructed with the required resolution limit.

A necessary criterion of coherence can be immediately formulated, without any regard to details of the hologram. Imagine that an absolutely coherent illuminating beam is moved during the exposure parallel to itself, so that a representative point of it, e.g. the mean paraxial focus, fills a circular disk with diameter d_c. But this is equivalent to moving the object within a disk of the same diameter, as only the relative position of beam and object matters for the physical shadow at infinity, and from such a 'wobbled' hologram we could at best reconstruct an image with a resolution limit d_c. Thus we obtain the necessary condition that the Gaussian or nominal diameter of the illuminating disk, d_c, must not exceed the Abbe limit d_A

$$d_c \leqslant d_A = \tfrac{1}{2}\lambda/\sin\gamma_m. \tag{44}$$

But we can show that this necessary condition is also sufficient, because it will produce holograms practically indistinguishable from one taken with an absolutely coherent beam, within a plate radius corresponding to the maximum angle γ_m. Express in equation (47) the wave-length by de Broglie's relation as

$$\lambda = h/p,$$

where p is the momentum associated with the wave. This relation is valid for photons as well as for electrons or any other particles. Interpreting $p\sin\gamma_m$ as the maximum transversal momentum p_t of the particles in the beam, we write (47) in the form

$$d_c 2p_t \leqslant h. \tag{44·1}$$

Confront this with Heisenberg's inequality

$$d_c' 2p_t' \geqslant h, \tag{45}$$

where d_c' is the maximum transversal uncertainty of position of particles in the beam in the Gaussian focal plane, and $2p_t'$ is the maximum uncertainty in the transversal momentum. Consider first the case that the beam is limited by a physical aperture in the plane considered, i.e. $d_c = d_c'$. Heisenberg's principle states that if the particles composing the beam are specified to the limit (45), they are indistinguishable, that is to say, they produce effects, such as interference fringes, which cannot be distinguished from one another by observation within the cone-angle corresponding to that value of p_c' which changes the inequality into an equality. Comparing (44·1) and (45) we see that if $d_c = d_c'$, we must have $p_t < p_t'$, thus the interference fringes inside the cone γ_m are *a fortiori* the same for all beam particles.

But if d_c is not a physical aperture, but the Gaussian image of one, formed by an optical system, the criterion still holds, because $d_c \sin\gamma_m$ is an invariant in Gaussian optics. If the criterion (44) were not sufficient, it would be possible to break through Heisenberg's principle by placing a suitable lens system in front of the physical aperture to produce observable differences in the fringe system, which would make the particles to some extent distinguishable.

These very general considerations are of course uncertain to a factor of the order unity. In order to obtain a more quantitative idea of the changes which are produced in the hologram by departure from absolute coherence, consider the simple case of illumination through a physical aperture of diameter d, and investigate its effect on the fringe system produced by a point object on the axis, at a distance z_0 from the aperture. Each point of the illuminating aperture produces a fringe system concentric with the axis which connects this point with the point object. These fringe systems are incoherent with one another, hence their intensities must be summed. At the edge of the hologram the angular spacing of two fringes is $\lambda/z_0 \sin\gamma_m$. Two fringe systems will just wipe out one another if they are displaced by half this amount. This will be the case if the spacing of the two point sources is $\frac{1}{2}\lambda/\sin\gamma_m$, which is just the Abbe limit d_A.

With Zernike (1948), we define the 'degree of coherence' D_c, as the range of intensity difference between maxima and minima in the fringe system at the marginal angle γ_m, divided by the corresponding quantity if the same light flux issues from a point source at the centre of the aperture. Assuming that the intensity variation in the fringe system is sinusoidal, one obtains

$$D_c = \iint \cos(\pi x/d_A)\,dx\,dy \Big/ \iint dx\,dy, \tag{46}$$

where the integration has to be carried out over the area of the illuminating aperture of diameter d. The integrand $\cos(\pi x/d_A)$ expresses the fact that two points spaced in the X-direction by d_A just oppose one another. The integration gives

$$D_c = J_0\left(\tfrac{1}{2}\pi\frac{d}{d_A}\right) + J_2\left(\tfrac{1}{2}\pi\frac{d}{d_A}\right), \tag{47}$$

where J_0 and J_2 are the Bessel functions of zero and second order. Some values are

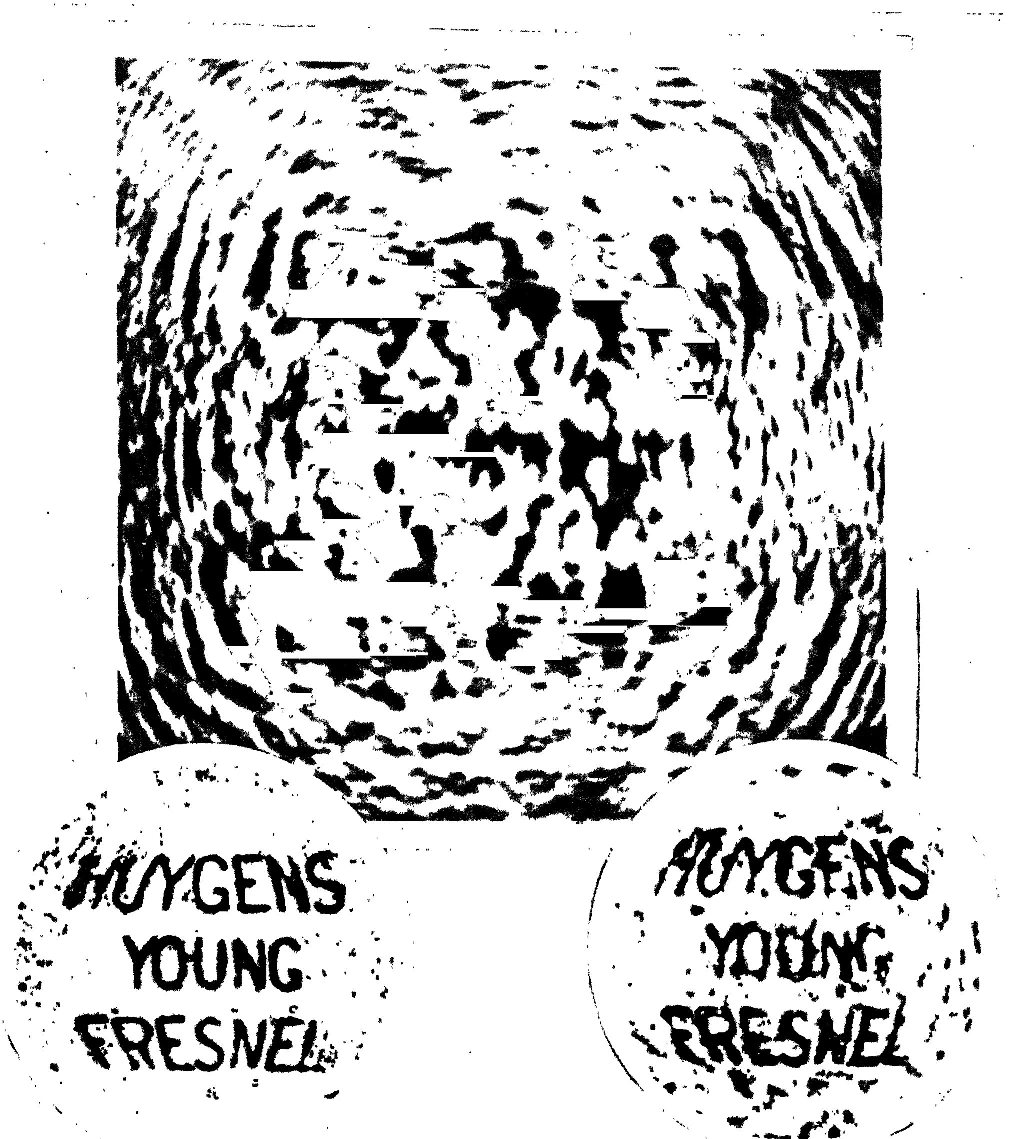
d/d_A	0	0·5	0·75	1·0	1·25	1·5	1·75	2·0
D_c	1	0·925	0·837	0·723	0·590	0·448	0·312	0·181

This justifies the expectation that the fringes system at the edge of the hologram will be rapidly effaced if the diameter of the light source appreciably exceeds the Abbe limit.

The coherence condition (47) represents a severe limitation of the available intensities, and it is the chief reason why the applications of the method of reconstructed wave-fronts will be probably restricted to light, with wave-lengths not very far from the visible, and to electrons. X-rays, protons and other particles will have to be excluded, as no sufficiently intense sources are available. Even in the case of electrons rather long exposures will be necessary, unless the present-day technique is improved.

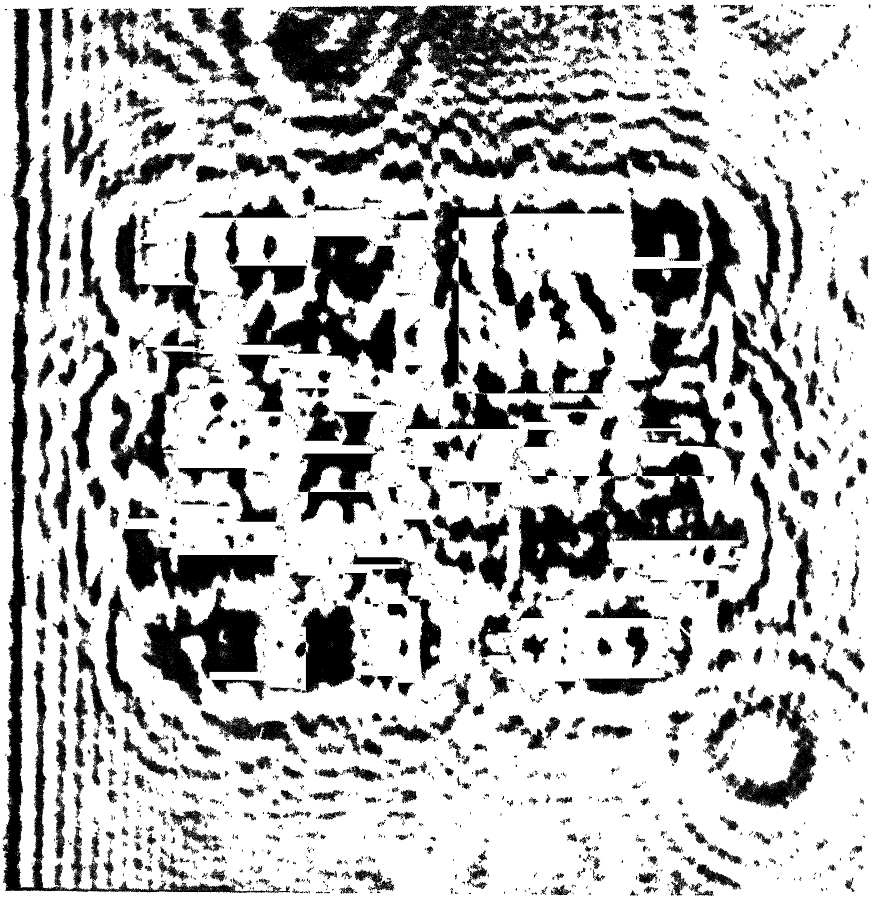
THE OPTICAL RECONSTRUCTION

So far we have assumed in the formulae, for simplicity, that the reconstruction is carried out with the same wave-length as used in the production of the diffraction pattern. Let us now distinguish the first wave-length by λ', the second by λ'', and use the primes $'$ and $''$ also for distinguishing the data A_s, C_s in the analyzer and in the synthetizer. The same formal distinction will be used also for z_0' and z_0'', but here a word of explanation is required. z_0' is a datum of the analysis; it is the actual distance of the object from the mean paraxial focus of the illuminating beam. But there is no

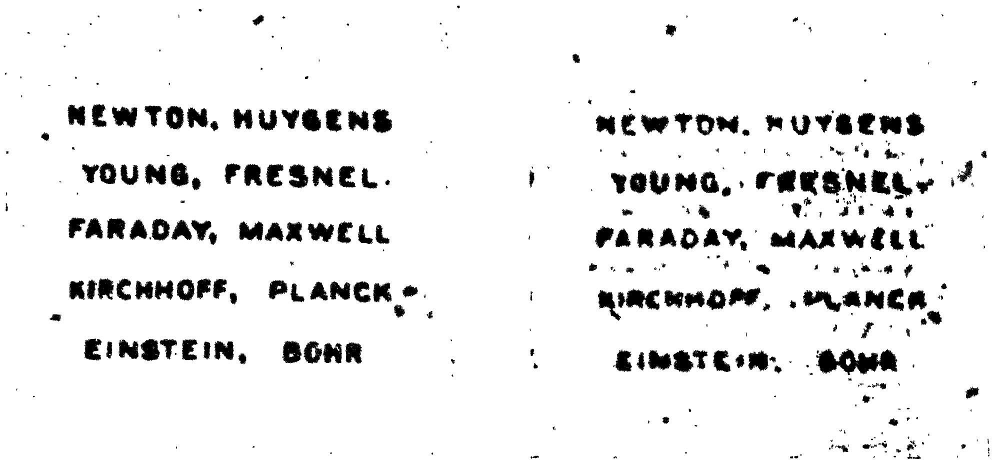


original hologram reconstruction

FIGURE 10. Optical reconstruction cycle. The original was a microphotograph of 1·5 mm. diam. Illuminated with $\lambda = 4358$ Å through pinhole 0·2 mm. diam., reduced by a microscope objective to 5μ nominal diameter, at 50 mm. from object. Geometrical magnification 12. Effective aperture of lens used in reconstruction 0·025. Noisy background chiefly due to imperfections of illuminating objective.



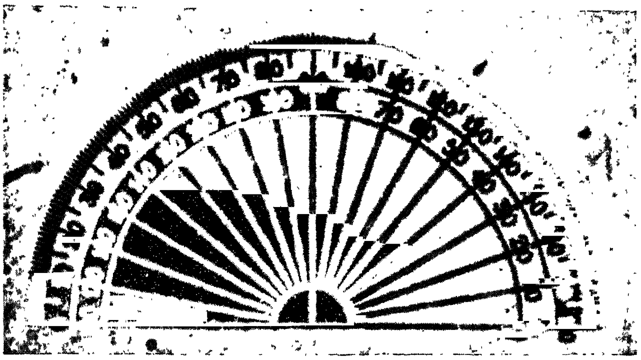
hologram



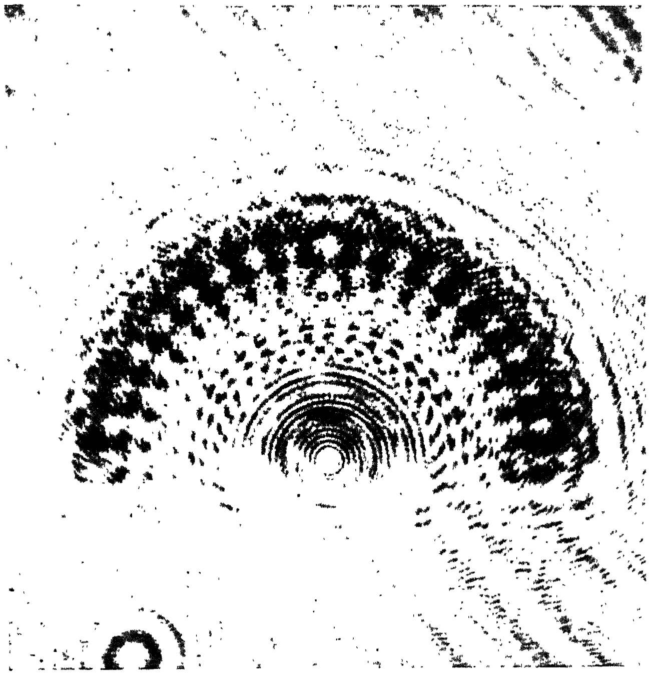
original

reconstruction

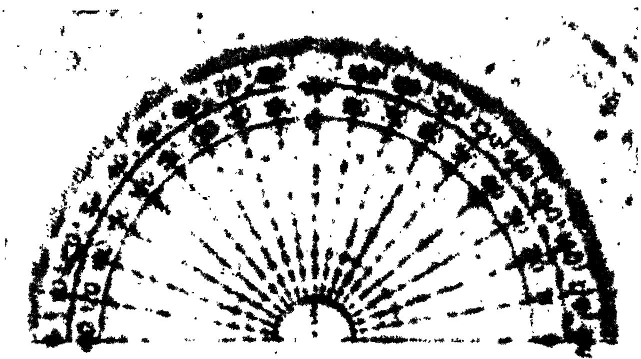
Figure 11. Reconstruction cycle with pinhole illumination. The letters in the original were inscribed in a rectangle $0{\cdot}65 \times 0{\cdot}5$ mm. Illumination with $\lambda = 4358$ Å through pinhole of 5μ diam. at 18 mm. from object. Geometrical magnification 10. Effective aperture used in reconstruction 0·075.



original



hologram



reconstruction

FIGURE 12. Reconstruction with pinhole illumination and wave-length change. The original was a micro-protractor of 1 mm. diam. Same conditions as in figure 11, but the wave-length used in the reconstruction was $\lambda = 5461$ Å.

physical object in the synthetizer, and z_0'' means merely the plane on which the viewing system must be focused in order to obtain a true, or at least the truest possible image of the original object.

The result of the analysis, the physical shadow, now to be called τ', is described by equation (42). We write down this equation again, but replace the Fourier variables ξ, η by the angles α, β, γ. For reasons of symmetry it will be convenient to attach the prime $'$ not only to the data of the analyzer, but also to the co-ordinates $x, y, \alpha, \beta, \gamma$ and $\alpha_0, \beta_0, \gamma_0$ used in the analysis. We write

$$\tau'(\alpha', \beta') = \iiiint \frac{A'(\alpha_0', \beta_0')}{A'(\alpha', \beta')} t'(x', y') \times \exp\{2\pi i[Q(\alpha_0', \beta_0') - Q(\alpha', \beta')]\} \frac{d(\cos\alpha_0')\, d(\cos\beta_0')}{\cos\gamma_0'} dx'\, dy', \tag{48}$$

where the phase function Q is

$$Q(\alpha, \beta') = \frac{1}{\lambda'}[x'\cos\alpha' + y'\cos\beta' - \tfrac{1}{2}z_0'\sin^2\gamma_0' - \tfrac{1}{2}A_s'(\cos^2\alpha' - \cos^2\beta') - \tfrac{1}{8}(z_0' + 2C_s')\sin^4\gamma']. \tag{49}$$

The same equation applies to the synthesis, i.e. to the reconstruction of an object t'', with all primes $'$ changed into $''$. The fact that the hologram obtained in the analysis is used in the reconstruction is expressed by

$$\tau''(\alpha'', \beta'', \gamma'') = \tau'(\alpha', \beta', \gamma'), \tag{50}$$

where the angles α', β', γ' and $\alpha'', \beta'', \gamma''$ belong to corresponding points of the hologram. The relation between them is given by the geometries of the analyzer and of the synthetizer.

Consider first the simple case, illustrated in figure 1, in which the focal length f of the collimator lens in the synthetizer, which moves the hologram optically to infinity, is the same as the throw L in the analyzer. In this case the angles α', β' and α'', β'' are the same, and their primes can be disregarded. It can be seen by inspection of equation (48) that it is transformed into the corresponding equation for $\tau'' = \tau'$ if we put

$$x' = \frac{\lambda'}{\lambda''}x'', \quad y' = \frac{\lambda'}{\lambda''}y'', \quad A_s' = \frac{\lambda'}{\lambda''}A_s'', \quad C_s' = \frac{\lambda'}{\lambda''}C_s'', \quad z_0' = \frac{\lambda'}{\lambda''}z_0'', \tag{51}$$

and

$$\tau''(x'', y'') = \tau'\left(\frac{\lambda'}{\lambda''}x'', \frac{\lambda'}{\lambda''}y''\right). \tag{52}$$

The transformation of the integration variables is purely formal. The next two equations postulate the scaling up of the aberrations A_s', C_s' in the synthetizer, and the last of the conditions (51) states that one must focus on the plane z_0'' in order to see the object t'' given by equation (52).

Consider now the more general case

$$f = kL, \tag{53}$$

i.e. we use a collimator lens of focal length k times the throw in the analyzer, always assuming of course that the hologram is in the focal plane of the lens. (This covers also the case in which the hologram used in the synthesis is an m times enlarged replica of the original; in this case the parameter k which figures in the following equations has the value f/mL.) The angles $\alpha' \dots$ and $\alpha'' \dots$ are now connected by the relations

$$\frac{\cos\alpha'}{\cos\gamma'} = k\frac{\cos\alpha''}{\cos\gamma''}, \quad \frac{\cos\beta'}{\cos\gamma'} = k\frac{\cos\beta''}{\cos\gamma''}. \tag{54}$$

The solution of these equations can be written in the form

$$\cos\alpha' = k\cos\alpha''[1-\tfrac{1}{2}(k^2-1)\sin^2\gamma''-\tfrac{3}{8}(k^2-1)^2\sin^4\gamma''-\dots]. \tag{54·1}$$

Only the first two terms of the expansion will be required. Introduce these into equation (48), where for simplicity we put $\lambda' = \lambda''$, to separate the change of geometry from the change of wave-length. The essential properties of the transformation can be deduced from the phase function Q, equation (49), which now assumes the form

$$\begin{aligned}\lambda Q = {} & k\,(x'\cos a''+y'\cos\beta'')-\tfrac{1}{2}k^2z_0'\sin^2\gamma''-\tfrac{1}{2}k^2A_s'(\cos{}^2\alpha''-\cos^2\beta'') \\ & -\tfrac{1}{8}k^4\left(\frac{z_0'}{k^2}+2C_s'\right)\sin^4\gamma''-\tfrac{1}{2}k(k^2-1)\,(x'\cos\alpha''+y'\cos\beta'')\sin^2\gamma'' \\ & +\tfrac{3}{8}z_0'k^2(k^2-1)\sin^4\gamma''+\tfrac{1}{2}A_s'k^2(k^2-1)\sin^2\gamma''\,(\cos^2\alpha''-\cos^2\beta'').\end{aligned} \tag{55}$$

The terms in the first row and the first term in the second correspond to an exact reproduction, the others represent errors which arise only if $k^2 \neq 1$. Considering the first four terms only, equation (48) transforms into an identical equation for τ'' instead of τ' by putting

$$kx' = x'', \quad ky' = y'', \quad k^2A_s' = A_s'', \quad k^4C_s' = C_s'', \quad k^2z_0' = z_0'', \tag{56}$$

and

$$t''(x'', y'') = t'\left(\frac{x''}{k}, \frac{y''}{k}\right). \tag{57}$$

This means that in order to see an image which is a k times enlarged replica of the original we must scale up the astigmatism k^2 times, the spherical aberration k^4 times, and focus the viewing system on a plane $z_0'' = k^2z_0'$.

But this image will appear with certain aberrations, which are indicated by the new terms in (55). The second term in the second row represents a *coma*. The first term in the last row is an addition to the spherical aberration, which can be incorporated in C_s''. The last term shows that the astigmatism A_s' of second order in the analyzer has produced astigmatism of the fourth order in the analyzer, i.e. a spherical aberration of the elliptical type.

All these error terms can be kept very small unless $k^2 \ll 1$. It can be shown that the best positions of the object are near $z_0' = -C_s'\sin^2\gamma_m'$, hence x', y' will be of the order $C_s'\sin^3\gamma_m'$, even if the object is in a marginal position. Hence the coma term in (55) will be of the order

$$k(k^2-1)\,C_s'\sin^3\gamma_m'\sin^3\gamma_m'' \simeq \frac{k^2-1}{k^2}\,C_s'\sin^6\gamma_m',$$

i.e. unless $k^2 \ll 1$ this will be a very small term, except in extreme cases when the spherical aberration $C_s'\sin^4\gamma_m'$ is of the order of several hundred fringes. In such cases the coma might amount to a few fringes, and coma compensation in the viewing system may become necessary.

The last term in (55) is of the order

$$\frac{k^2-1}{k^2} A'_s \sin^4 \gamma'_m,$$

which is again very small unless $k^2 \ll 1$. A'_s in good electron lenses is 10^{-4} or even less of C'_s; thus even if the spherical aberration is of the order of a thousand fringes, this term will represent a fraction of a fringe only.

Thus it is admissible to make the length of the optical synthetizer appreciably different from the throw in the electronic analyzer. It may be particularly advantageous to make $k < 1$, that is to say, not to make use of the full magnification λ''/λ' which is about 100,000, but only of a part of it. The rest can be supplied by the viewing system. This has the advantage that one can work with smaller lenses, though with proportionately larger numerical aperture. Assuming, for instance, $C'_s = 1$ cm. and $\sin \gamma'_m = 0{\cdot}05$, the minimum diameter of the electron beam is $0{\cdot}625\,\mu$, and if one makes $k = 1$ one requires an optical system capable of handling a light beam with 6·25 cm. minimum diameter. It will be advantageous to reduce this to one-half, or even to one-quarter, as optical systems with numerical apertures of 0·1 to 0·2 present no difficulties if the lenses need not be large.

To sum up, if in the optical synthetizer the data of the electronic condenser system are scaled up according to

$$A''_s = k^2 \frac{\lambda''}{\lambda'} A'_s, \quad C''_s = k^4 \frac{\lambda''}{\lambda'} C'_s, \tag{58}$$

the transversal dimensions of the object will appear scaled up in a ratio $k\lambda''/\lambda'$ and the longitudinal dimensions in the ratio $k^2\lambda''/\lambda'$. Thus *the geometrical or k-part of the transformation is of the type as produced by optical instruments, with a longitudinal magnification equal to the square of the transversal, while the λ-part is a uniform scaling-up, not realizable by ordinary optical imagery.*

The accuracy with which the conditions (58) have to be fulfilled can be best stated in terms of fringes. The maximum admissible deviation of a wave-front from the spherical shape without loss of resolving power has been estimated by Glaser (1943) as 0·4 of a wave-length, by Bruck (1947) as one wave-length. The second can be considered as the more reliable estimate. Thus the condition (58) for C''_s must be observed to an accuracy of one fringe. Assuming again $C'_s = 1$ cm. and a resolution limit of 1 Å, one requires by Abbe's rule an aperture $\sin \gamma'_m = 0{\cdot}025$, and with the more accurate numerical factor 0·6, $\sin \gamma'_m = 0{\cdot}030$. This gives 200 or 400 fringes at the edge of the field, according to which numerical factor one adopts. Thus the spherical aberration in the optical model must imitate C'_s to about one fringe in 200 or in 400.

The astigmatism tolerance at the edge of the field is about a quarter fringe. In carefully manufactured electron objectives A_s is of the order of a few microns, and it can be reduced by the compensation methods introduced by Hillier & Ramberg (1947) by at least one order of magnitude. This is necessary for realizing the full resolving power of present-day electron microscopes. In terms of fringes, the astigmatism in carefully manufactured but not compensated electron lenses amounts to a few fringes at apertures of 0·003, and if this is opened up ten times, to realize a ten times improved resolving power, the distortion will be of the order of a few hundred fringes. Thus A'_s must be also imitated in the optical synthetizer to an accuracy of one part in a few hundred.

One could think of imitating the data of the electron-optical system by first carefully measuring A'_s and C'_s and computing an optical system with these data. But this is hardly a practicable method. Apart from the difficulties of measuring to the required accuracy, by the time the computation is finished and the optical replica is made the data of the electron-optical system are likely to have changed by far more than the error tolerance. It will be much preferable to make the astigmatism and the spherical aberration of the synthetizer variable, and adjust them until certain known parts of the object, such as the support, or certain standard test objects appear with maximum sharpness. The spherical aberration can be made variable by shifting a fourth-order plate, the astigmatism by crossed cylindrical lenses or by tilting lenses. Expert opticians will be doubtlessly able to work out a schedule to carry out the three adjustments of focus, astigmatism and spherical aberration in a systematic way. Thus only a moderate degree of constancy is required of the electron-optical system, sufficient at least for a series of reconstructions, without too frequent readjustments.

EXPERIMENTAL TESTS

Experiments were started almost as soon as the idea of reconstruction first emerged. They confirmed the soundness of the basic principle, but pointed to the necessity of elaborating and modifying the original, somewhat primitive views on the mechanism of reconstruction, which have been described elsewhere (Gabor 1948). The experiments were later continued in order to test the conclusions from the quantitative theory described in this paper.

In these tests analysis and synthesis were both carried out with visible light, though not always with the same wave-length. The arrangement for taking holograms was substantially as shown in the upper part of figure 1, but with optical instead of with electron lenses. A condenser threw an image of a high-pressure mercury arc (of the 'compact' type, with tungsten electrodes) through a colour filter on an aperture of about 0·2 mm. diameter. The lines used were 4358Å (violet), and 5461Å (green), isolated by Wratten light filters nos. 47 and 61. In the earlier tests a microscope objective was used to produce an image of this aperture, about 40 times reduced, i.e. with a nominal diameter of about 5μ, which formed the 'point source'. The objects were mostly microphotographs, sandwiched with immersion oil between two polished glass plates. In the earlier experiments the distance between the point source and the object was about 50 mm., the distance from the object to the photographic plate 550 mm., thus the geometrical magnification was about 12.

The photographic plate was held in position against three locating pins. Originally it was planned to develop the holograms by reversal, to make sure of exactly identical positions in the analysis and in the synthesis. In the negative-positive process the printing was carried out on the same locating pins. These precautions proved unnecessary in those experiments in which not only the Gaussian but also the physical diameter of the source was of the order of the resolution limit, which proves that in these cases the theory of homocentric illuminating beams is a satisfactory approximation. But they were required later, in experiments with very strong spherical aberration in the illuminating beam. Reversal development, however, was found unnecessary, and the far more flexible negative-positive photographic process was used throughout. The negative hologram was usually processed with $\Gamma = 1{\cdot}2$ to $1{\cdot}6$, and the positive with $\Gamma = 0{\cdot}7$ to $1{\cdot}6$, so that a wide range of overall gammas could be tested. When it was confirmed that an overall gamma of 2 gave the best results, this was realized as closely as possible.

In the reconstruction the positive hologram was sandwiched with immersion oil between polished glass plates, which had to be carefully selected. It was backed by a viewing lens, which was an achromatic doublet, cemented and bloomed, with a focal length of 175 mm. and a linear aperture of 47 mm. The spherical aberration was 3 fringes at infinite conjugates. The diameter which satisfies the quarter-wave tolerance can be estimated at 27 mm., and the numerical aperture figures given below are based on this 'effective diameter'. The reconstructed image was viewed in a microscope, and photographed on plates introduced into the eyepiece.

Figure 10, plate 15, is a record of one of these earlier experiments. The figure at the left is a direct photograph of the original, which was a microphotograph of the names of the three founders of the wave theory of light. It was taken through the viewing system, with the same optics as used for the reconstruction. The top figure is the central part of the hologram, and the one at the right, is the reconstruction. All three were taken with the violet mercury line 4358 Å. The effective numerical aperture was 0·025, thus the resolution limit $0{\cdot}6 \times 0{\cdot}436/0{\cdot}025 = 10\mu$. This is $\frac{1}{150}$ of the diameter of the reproduced part of the microphotographs, and corresponds about to the gap between the 'Y' and the 'G' in 'HUYGENS'.

Though in its best parts the reconstruction almost attains the resolution of the direct photograph, the picture is very 'noisy'. This is due only to a smaller part to the essential disturbance created by the twin image, to a greater part it is due to specks of dust, and inhomogeneities in the two microscope objectives. It may be noted that these very troublesome effects, unwelcome concomitants of the great phase-discriminating power of the methods using a coherent background, cannot be expected to appear in an electronic analyzer. However imperfect an electron lens may be from the point of view of theoretical optics, it can contain neither dust nor 'schlieren', as the electromagnetic field smoothes itself out automatically, and in this respect any electron lens is superior to all but the best optical lenses.

In order to avoid these inessential disturbances, in some later experiments the optical surfaces were reduced to a minimum. In the experiments of which figures 11 and 12, plates 16 and 17, are records, the source was a pinhole of 3μ diameter, pierced into tinfoil with a very fine needle. Thus no glass surfaces other than those of the microphotographs were involved in the taking of the hologram. In the reconstruction the optics was also reduced to a minimum by cutting out the second microscope. The spacing between the object and the viewing lens was reduced to 180 mm., the distance between the lens and the plate increased to 700 mm., so that a fourfold enlargement of the object was produced by the viewing lens, sufficient for direct photography on not unduly slow plates. Further enlargement was obtained in some cases by taking the hologram with the violet line, but reconstructing it with the green line.

The effective numerical aperture in this experimental series was 0·075, and the theoretical resolution limit 3·5μ. This is about $\frac{1}{350}$ of the diameter of the part of the microphotograph which is reproduced in figure 11 and which contains ten great names in the theory of light. The resolution is just about sufficient to resolve the hole in an 'A'. The theoretical resolution of the reconstruction is less, because the pinhole source of 3μ, used both in the analysis and in the synthesis, is of the same order. It can be estimated at about 5·5μ, by the thumb rule of orthogonal composition of errors. This resolution has been in fact very nearly achieved in the case of figure 11 and also in figure 12. It can be also seen that the background is very much more even than in figure 10. The residual disturbance is mostly essential, and due to the twin object. In these experiments the twin object could be separately focused, and as regards sharpness could not be distinguished from the 'true' image.

Experiments for testing the theory in the case of illuminating beams with large spherical aberration are in progress, but they have already confirmed its main results.

Conclusion

The new principle can be applied in all cases where coherent monochromatic radiation of sufficient intensity is available to produce a divergent diffraction pattern, with a relatively strong coherent background. While the application to electron microscopy promises the direct resolution of structures which are outside the range of ordinary electron microscopes, probably the most interesting feature of the new method for light-optical applications is the possibility of recording in one photograph the data of three-dimensional objects. In the reconstruction one plane after the other can be focused, as if the object were in position, though the disturbing effect of the parts of the object outside the sharply focused plane is stronger in coherent light than in incoherent illumination. But it is very likely that in light optics, where beam splitters are available, methods can be found for providing the coherent background which will allow better separation of object planes, and more effective elimination of the effects of the 'twin wave' than the simple arrangements which have been investigated.

I thank Mr L. J. Davies, Director of Research of the British Thomson-Houston Company, for permission to publish this paper and Mr J. Williams for assistance in the experimental work.

References

Baker, B. B. & Copson, E. T. 1939 *The mathematical theory of Huygens' principle.* Oxford: Clarendon Press.

Boersch, H. 1938 *Z. techn. Phys.* **19**, 337.

Boersch, H. 1939 *Z. techn. Phys.* **20**, 346.

Bragg, W. L. 1942 *Nature*, **149**, 470.

Bruck, H. 1947 *C.R. Acad. Sci., Paris*, **224**, 1553.

Bunn, W. 1945 *Chemical crystallography.* Oxford University Press.
Campbell, G. A. & Foster, R. M. 1931 *Fourier integrals for practical applications.* Bell Telephone System Monograph, B. 584. New York.
Gabor, D. 1948 *Nature,* **161**, 777.
Glaser, W. 1943 *Z. Phys.* **121**, 647.
Hillier, J. & Ramberg, E. G. 1947 *J. Appl. Phys.* **18**, 48.
Zernike, F. 1948 *Proc. Phys. Soc.* **61**, 147.

The thermal equilibrium at the tropopause and the temperature of the lower stratosphere

By R. M. Goody, *St John's College, University of Cambridge*

(*Communicated by D. Brunt, F.R.S.—Received* 16 *August* 1948—
Revised 14 *February* 1949)

Further considerations along the lines of Emden's methods lead to continuity of temperature at the tropopause as a condition for a stable transition from a state of convective to radiative equilibrium. This explains the characteristic appearance of the temperature distribution near the tropopause. Application of this condition leads to a simple explanation of the latitude variation of stratosphere temperature, mainly in terms of the effects of water vapour and carbon dioxide. The variation of stratosphere temperature with ozone concentration may be calculated, which confirms Dobson's hypothesis that anomalous seasonal variations in stratosphere temperature are due to seasonal variations of ozone concentration. Reasons for the approximately isothermal character of the lower stratosphere are also discussed.

1. INTRODUCTION

The facts relating to the temperature of the lower stratosphere which require explanation are now well established. They are:

(i) that the transition of temperature from the troposphere to the stratosphere is smooth;

(ii) that the temperature is lower over the tropics than over the arctic;

(iii) that temperature gradients are small relative to those occurring in the troposphere, and often positive; and

(iv) that the seasonal variation of temperature differs from that of the troposphere immediately below the tropopause.

The lower stratosphere is taken to mean the approximately isothermal region between the tropopause and a height of approximately 30 km., above which temperature begins to increase rapidly with height.

Any theory of heat interchange in the lower stratosphere must attempt to explain these facts. Such a theory, in the present state of knowledge, must necessarily be highly simplified, for the existence of winds at high altitudes, and of moving air masses with distinct boundaries, make an exact treatment almost impossible. In problems of similar complication, however, it is sometimes possible to isolate one feature and to deduce a restrictive condition which requires relations similar to those observed in nature. If this can be done, then a framework has been erected upon which it may be possible to build a more complete theory.

This paper will only consider the thermal equilibrium at the tropopause upon the assumption that it represents the point of division between a region where heat exchange is mainly due to convection (the troposphere), and a region where heat exchange is due to radiation (the stratosphere). This assumption was made by Gold (1909) in his pioneer work on the stratosphere, and it is still generally accepted for the following reasons.

The troposphere is perpetually stirred by winds, and even if winds were not present Emden (1913) has shown that irradiation of the earth by the sun would lead to convective instability in the atmosphere's lower layers, and require them to be in convective equilibrium. On the other hand, the lower stratosphere is approximately isothermal, and by Richardson's criterion the wind gradients required to produce atmospheric instability are nearly twice as great in the lower stratosphere as in the upper troposphere. Thus turbulent transfer of heat may reasonably be expected to be of minor importance in the lower stratosphere. Since states of convective and radiative equilibrium both require continuity of temperature gradient, provided that there is continuity of material density, the discontinuity of temperature gradient at the tropopause indicates a change in the mode of heat transfer. Heat exchanges may therefore plausibly be taken to be dominated by convection and turbulence in the troposphere, and radiation in the stratosphere. The beginning of partial diffusive separation of oxygen and helium approximately at the tropopause, as observed by Regener (1936) and Paneth & Glückauf (1935), lends weight to this assumption.

This paper is only concerned with the troposphere through the upward flux of radiation which it supplies at the tropopause. The temperature lapse rate in the troposphere is taken as 6·5° K/km. independent of season and latitude, as observations indicate. The factors leading to this lapse rate are not considered.

In order to treat the problem as an equilibrium problem it is necessary to neglect heating by direct solar radiation at the tropopause, and the movement of air masses. Other authors (for example, Dobson 1946) have argued that the direct heating by solar radiation at the tropopause is negligibly small compared with the heating by diffuse terrestrial radiation. Movement of air masses constitutes a more serious difficulty, and it must be assumed that the observed average conditions at a fixed point on the earth's surface correspond to the equilibrium conditions at that point for a static atmosphere.

Dobson's paper (1946) on the temperature of the lower stratosphere considers empirically the effect of the presence of several different gases each with a complex absorption spectrum. The present paper is an attempt to find a secure theoretical basis for such calculations. The line of attack is similar to that adopted by Emden, but in carrying it through to a logical conclusion a criterion for stability at the tropopause is found which is essentially similar to that adopted by Gold. This synthesis of the work of Gold and Emden is a happy development, for despite their apparent incompatibility both theories seemed to touch upon the truth.

2. SYMBOLS

Following Schwarzschild (1906) the terrestrial radiation will be supposed to be confined to two anti-parallel streams. According to Elsasser (1942) this assumption is justified for diffuse radiation where the absorbing gases possess band spectra, provided an absorption coefficient 1·66 times that appropriate for parallel radiation is employed. Since absorption coefficients of atmospheric gases are not known with great accuracy, this simplification will hardly impair the precision of the results and will not affect such features as arise from the presence of several different absorbing gases each with different spatial distributions and absorption spectra.

Symbols are defined as follows:

x, y, z	vertical height.
$\theta(x)$	temperature at x. It will often be used as an alternative variable to x, thus $B_\lambda(\theta) = B_\lambda(\theta(x)) = B_\lambda(x)$.
λ	wave-length.
$B_\lambda(x)\,\Delta\lambda$	total energy radiated per cm.^{2}/sec. in the wave-length range λ to $\lambda+\Delta\lambda$ from a black body at temperature $\theta(x)$.
$\alpha_\lambda(x)\,\Delta\lambda$	upward flux of radiation per cm.^{2}/sec. in the wave-length range λ to $\lambda+\Delta\lambda$ at x.
$\beta_\lambda(x)\,\Delta\lambda$	downward flux of radiation per cm.^{2}/sec. in the wave-length range λ to $\lambda+\Delta\lambda$ at x.
$\chi_\lambda(x) = \alpha_\lambda(x)+\beta_\lambda(x)$	total radiation at x.
$\phi_\lambda(x) = \alpha_\lambda(x)-\beta_\lambda(x)$	flux of radiation at x.
$k_\lambda(x)$	mass absorption coefficient of the absorbing medium at the point x at wave-length λ.
$\rho(x)$	density of absorbing matter at the point x.

In addition, s and t will be used as the heights of two points an infinitesimally small distance apart, above and below the tropopause respectively. For brevity functions like $(\partial B_\lambda(x)/\partial x)_{x=s}$ will be written $\partial B_\lambda(s)/\partial x$; no confusion should arise.

The parallel radiation assumption leads to the following forms for the equations of radiative equilibrium and transfer (see, for example, Hopf 1934):

$$\int_0^\infty \{2B_\lambda(x) - \chi_\lambda(x)\}\, k_\lambda(x)\,\rho(x)\,d\lambda = 0, \tag{1}$$

$$\frac{\partial \chi_\lambda(x)}{\partial x} = -k_\lambda(x)\,\rho(x)\,\phi_\lambda(x). \tag{2}$$

3. THE CONDITION FOR A STABLE CHANGE-OVER FROM CONVECTIVE TO RADIATIVE EQUILIBRIUM

(a) *General discussion*

Emden showed that if the earth's atmosphere were allowed to reach a state of radiative equilibrium with the solar radiation, the lower layers would have a super-adiabatic lapse rate of temperature and would be hydrodynamically unstable. Convection would therefore commence and spread upwards from the earth's surface. Emden did not consider how far this convective region would rise, but was content with explaining why such a region should occur. Calculations of the extent of the convective region are complicated by the fact that temperature discontinuities are possible at the boundary of a region of radiative equilibrium. The convective region will clearly rise as long as there is an unstable lapse rate in the radiative region. However, it will be shown that when the lapse rate in the radiative region is just stable, an unstable negative temperature discontinuity may be expected at the boundary, which would cause convection to continue in this boundary layer, even if it does not in the body of the radiative region. It may be shown, moreover, that when convection has ironed out the unstable discontinuity, then the lapse rate in the radiative region may be expected to be stable. The absence of temperature discontinuity at the boundary may therefore be expected to be the stable state to which the earth's atmosphere will tend.

The frequently observed absence of temperature discontinuities at the tropopause is usually accepted as commonplace and scarcely worthy of explanation. The subsequent analysis shows, however, that this condition must be achieved if the troposphere and stratosphere are to be in equilibrium, and the common absence of a marked temperature discontinuity shows that equilibrium or near equilibrium frequently occurs.

Real temperature discontinuities naturally cannot exist, for they would imply infinitely large heat transport by molecular processes. In the earth's atmosphere such discontinuities will really be shallow regions of high gradients of temperature. If the lapse rate in such a region is greater than adiabatic, no matter how shallow the region, contact with the convective troposphere below will cause it to be disturbed.

In § 3 (*b*) two theorems will be proved for a 'grey' absorbing atmosphere. These theorems will be used in § 3 (*c*) to elaborate the argument that the absence of temperature discontinuity at the tropopause is a sufficient condition for a stable atmosphere. In § 3 (*d*) the differences between the earth's atmosphere and a 'grey' atmosphere are considered in order to show that the differences lead to a tightening up of the arguments of § 3 (*c*).

(*b*) *Two theorems*

(i) A grey absorbing atmosphere ($k_\lambda = k$, const.) will be considered, for which it is not necessary to employ λ suffixes. Let

$$\alpha(x) = \int_0^\infty \alpha_\lambda(x)\, d\lambda,$$

$$B(x) = \int_0^\infty B_\lambda(x)\, d\lambda = \sigma\theta(x)^4,$$

$$\chi(x) = \int_0^\infty \chi_\lambda(x)\, d\lambda,$$

$$\phi(x) = \int_0^\infty \phi_\lambda(x)\, d\lambda = \text{const. in the stratosphere.}$$

Equations (1) and (2) may now be written

$$2B(x) = \chi(x), \quad \frac{\partial\chi(x)}{\partial x} = -k(x)\rho(x)\phi.$$

Since $k(s)\,\rho(s) = k(t)\,\rho(t)$, these equations may be put together in the form

$$\frac{\partial B(s)}{\partial x}+k(t)\,\rho(t)\,(\alpha(t)-B(t)) = k(s)\,\rho(s)\,(B(s)-B(t)). \tag{3}$$

(ii) A limit is required for the difference $\alpha(t)-B(t)$. The upward flux of radiation at the tropopause $\alpha(t)$ is given by Beer's law

$$\alpha(t) = B(0)\exp\left[-\int_0^t k(y)\,\rho(y)\,dy\right]+\int_0^t k(x)\,\rho(x)\,B(x)\exp\left[-\int_x^t k(y)\,\rho(y)\,dy\right]dx.$$

By partial integration this gives

$$\alpha(t)-B(t) = -\int_0^t \frac{\partial B(x)}{\partial x}\exp\left[-\int_x^t k(y)\,\rho(y)\,dy\right]dx.$$

In the troposphere $\partial\theta(x)/\partial x$ is constant, which means that $\partial B(x)/\partial x \propto \theta(x)^3$. If the absorbing gases are mixed in equal proportion with the atmosphere at all heights in the troposphere, then $\rho(y) \propto \theta(y)^{4\cdot 26}$. A true grey absorber will have a coefficient independent of pressure, but if it be assumed that $k(x)$ varies as some fractional power of $p(x)$ then $\rho(y)\,k(y)$ will vary more rapidly than $\theta(y)^{4\cdot 26}$. As long as $\rho(y)\,k(y)$ increases more rapidly than $\partial B(x)/\partial x$ the following inequality holds:

$$\alpha(t)-B(t) < -\int_0^t \frac{\partial B(t)}{\partial x}\exp\left[-\int_x^t k(t)\,\rho(t)\,dy\right]dx,$$

or

$$k(t)\,\rho(t)\,(\alpha(t)-B(t)) < -\frac{\partial B(t)}{\partial x}. \tag{4}$$

The difference will be the greater the more rapidly $\rho(y)\,k(y)$ varies with $\theta(y)$.

(*c*) *The transition from convective to radiative equilibrium in a grey absorbing atmosphere*

From Emden's description of the convective layer in the atmosphere it is apparent that this layer will extend upwards at least until the radiative region is hydrodynamically stable. Since the maximum stable lapse rate is affected by wind gradients, it is best to assume that in the troposphere turbulence neither increases nor decreases, and hence that on the average $-\partial\theta(s)/\partial x < -\partial\theta(t)/\partial x$ is the criterion for hydrodynamic stability. Suppose that the tropopause has risen until the lapse rate in the radiative region has dropped to this limiting value, then (3) and (4) give

$$k(s)\,\rho(s)\,\theta(s)^4-4\theta(s)^3\frac{\partial\theta(t)}{\partial x} < k(s)\,\rho(s)\,\theta(t)^4-4\theta(t)^3\frac{\partial\theta(t)}{\partial x}.$$

Since $\partial\theta(t)/\partial x$ is negative, this relation can only be satisfied if $\theta(s) < \theta(t)$, i.e. if there is a negative discontinuity of temperature at the tropopause.

This corresponds to the state shown in figure 1*a*. This state, while just stable in the lower stratosphere, is, nevertheless, unstable in a shallow layer just above the tropopause, which will be disturbed, and the tropopause will rise. Clearly the tropopause will continue to rise as long as this unstable layer exists. When the discontinuity has been ironed out a configuration corresponding to figure 1*b* will have been reached. Here $\theta(s) = \theta(t)$, and from (3) and (4)

$$-\frac{\partial B(s)}{\partial x} < -\frac{\partial B(t)}{\partial x} \quad \text{or} \quad -\frac{\partial \theta(s)}{\partial x} < -\frac{\partial \theta(t)}{\partial x}.$$

This means that provided the unstable layer is missing the body of the stratosphere will be hydrodynamically stable. The state 1*b* is therefore completely stable. It only remains to consider the state 1*c*, where there is a positive temperature discontinuity at the tropopause. This state also may appear to be stable, since by extrapolation from the states 1*a* and 1*b* it may be expected that $-\partial\theta(s)/\partial x \ll -\partial\theta(t)/\partial x$. If, however, convection were to cease just below the tropopause, a lapse rate would result which would be more stable than the lapse rate with convection in progress. This may reasonably be taken as a criterion for the cessation of turbulence just below the tropopause, in which case the tropopause will drop and once more the state 1*b* will be reached.

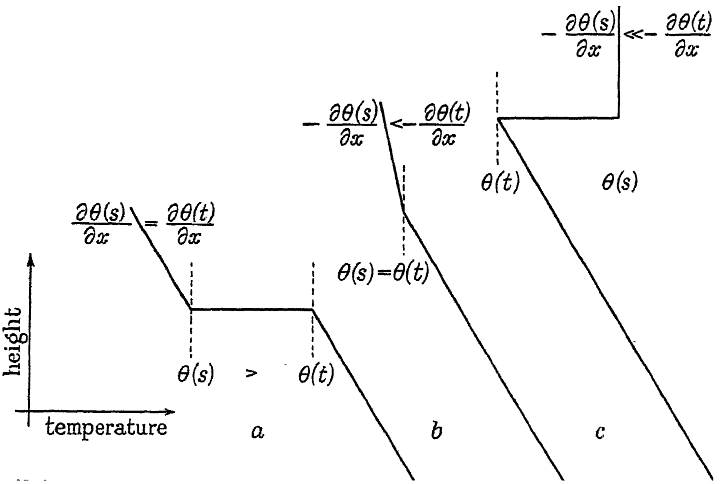

$-\frac{\partial\theta(s)}{\partial x} \ll -\frac{\partial\theta(t)}{\partial x}$
$-\frac{\partial\theta(s)}{\partial x} < -\frac{\partial\theta(t)}{\partial x}$
$\theta(t)$
$\theta(s)$
$\frac{\partial\theta(s)}{\partial x} = \frac{\partial\theta(t)}{\partial x}$
$\theta(s)=\theta(t)$
height
$\theta(s)$ > $\theta(t)$
temperature
a
b
c

FIGURE 1. Possible temperature distributions in the atmosphere.

(*d*) *The transition in the earth's atmosphere*

The main differences between the model adopted in § 3 (*b*) and the earth's atmosphere arise from the following factors:

(i) Absorption by atmospheric gases only takes place over finite band widths.

(ii) Transmission functions over these finite band widths differ greatly from the exponential Beer's law.

(iii) At least three important absorbing gases are present.

(iv) Of these three only carbon dioxide is mixed with the main oxygen-nitrogen mixture at all heights.

The whole argument of § 3 (*c*) is tightened up if the inequality (4) is increased. The factors (i) and (iv) tend to do this. As will be seen in § 4 the really important absorption bands of water vapour and carbon dioxide occur on the long wave-length side of the maximum emission from a 200 to 300° K black body. This means that $B_\lambda(x)$ for these bands will not vary as rapidly as $\theta(x)^4$, which increases the inequality (4). It will be seen later (see figure 2) that for most latitudes water vapour is the most important atmospheric gas other than carbon dioxide. The density of water vapour varies more rapidly than $\theta(x)^{4\cdot 26}$, which again increases the inequality (4). (ii) will not greatly affect the arguments of § 3 (*c*), since equations similar to (3) and (4) may be obtained using non-exponential transmission functions. The importance of (iii) is only clear if considered with the results of § 7, where it is shown that it is characteristic of a mixture of absorbing gases, that the lapse rate in the radiative region is less than it would be if only one absorbing gas were present. This means that the state 1*b* would have an even more stable lapse rate in the radiative region, and the state 1*a* would require a larger discontinuity to make $\partial\theta(s)/\partial x = \partial\theta(t)/\partial x$.

A full analysis of the processes occurring in the earth's atmosphere would probably be impossible, even if sufficiently accurate data were available, if only on account of the prohibitively laborious successive approximation calculations which would be required. The discussion given above goes to show, as far as is possible by simple arguments, that the processes described in § 3 (*c*) for a grey atmosphere will also occur in the earth's atmosphere. The calculations of § 7, which show a stable lapse rate in the lower stratosphere at all latitudes provided $\theta(s) = \theta(t)$, are some confirmation that detailed analysis will demonstrate the existence of the state 1*b*, which, according to the conclusions of § 5, must be stable.

4. APPLICATION OF THE EQUILIBRIUM CRITERION TO THE EARTH'S ATMOSPHERE

(*a*) *Plan of computation*

The equilibrium criterion discussed in § 3 may be applied in the following way. From a ground temperature $\theta(0)$, the temperature of the troposphere will fall at a rate of 6·5° K/km. to a value $\theta(t)$ at the tropopause. The stratosphere may be pictured as starting from the same temperature. Expressions may be written down for the rate of heating per cm.^{3} in a layer of the atmosphere just above the tropopause. When the height of the tropopause is such that there is no heating or cooling in this layer, a state of stable equilibrium will have been reached. In order to put such a calculation into a reasonably simple form, many approximations have to be made.

Suppose that there is no overlap between atmospheric absorption bands. This is a reasonable supposition except for the overlap between the 15μ carbon dioxide band and the 14·1μ ozone band, which will not lead to serious error. In the case of no overlap the absorption due to each band of each gas can be considered separately. Let $\int_r B_\lambda(\theta)\,d\lambda = B_r(\theta)$, etc., where the integration is over the limits of the *r*th absorption band, and let it be assumed that $B_\lambda(\theta)$ and $\partial B_\lambda(\theta)/\partial\theta$ do not vary greatly over each band.

Consider the two functions

$$T_r = \int_r (\alpha_\lambda(t) - B_\lambda(t))\,\rho(t)\,k_\lambda(t)\,d\lambda, \quad S_r = \int_r (\beta_\lambda(s) - B_\lambda(s))\,\rho(s)\,k_\lambda(s)\,d\lambda.$$

T_r is a function only of conditions in the troposphere, and represents the energy interchange by radiation between the lower surface of an infinitesimal layer just below the tropopause and the troposphere. S_r is similarly a function only of conditions in the stratosphere and gives the energy interchange between the upper surface of a layer just above the tropopause and the stratosphere.

If $B_\lambda(t) = B_\lambda(s)$, i.e. $\theta(t) = \theta(s)$, then the equation

$$\Sigma_r(T_r + S_r) = \Sigma_r H_r = 0 \tag{5}$$

is, from (1), the equation of radiative equilibrium for a point just above the tropopause. Thus equation (5) sums up the condition that there shall be radiative equilibrium in the lower stratosphere, with no temperature discontinuity at the tropopause.

By single partial integrations precisely similar to that used in § 3 (*b*)

$$\alpha_\lambda(t) - B_\lambda(t) = -\int_0^s \frac{\partial B_\lambda(x)}{\partial x} \exp\left[-\int_x^s k_\lambda(y)\,\rho(y)\,dy\right] dx,$$

$$\beta_\lambda(s) - B_\lambda(s) = -B_\lambda(\infty)\exp\left[-\int_s^\infty k_\lambda(y)\,\rho(y)\,dy\right] + \int_s^\infty \frac{\partial B_\lambda(x)}{\partial x} \exp\left[-\int_s^x k_\lambda(y)\,\rho(y)\,dy\right] dx.$$

The definitions of T_r and S_r then lead to

$$T_r = \int_0^s \frac{\partial B_r(x)}{\partial x} \frac{\partial \eta_r(x:s)}{\partial s}\,dx, \tag{6}$$

$$S_r = -B_r(\infty)\frac{\partial \eta_r(s:\infty)}{\partial s} + \int_s^\infty \frac{\partial B_r(x)}{\partial x}\frac{\partial \eta_r(s:x)}{\partial s}\,dx, \tag{7}$$

where η_r is the fractional transmission function of the rth band defined by

$$\eta_r(x:y) = \frac{\displaystyle\int_r d\lambda \exp\left[-\int_x^y \rho(z)\,k_\lambda(z)\,dz\right]}{\displaystyle\int_r d\lambda}.$$

Equation (6) may be evaluated without difficulty, since the temperature gradient in the troposphere is assigned, and the distribution of absorbing gases is fairly well known. Equation (7) is more difficult, since the distribution of temperature in the stratosphere cannot be assumed *a priori* but must be calculated by successive approximations, and knowledge of the distribution of water vapour in the stratosphere is limited to the lowest few kilometres. A way out of these difficulties may be found by appealing to observation, and setting $\partial\theta(x)/\partial x = 0$ in the stratosphere. This reduces (7) to

$$S_r = -B_r(s)\frac{\partial \eta_r(s:\infty)}{\partial s}, \tag{8}$$

which involves only the total absorbing power of the whole stratosphere, and is not affected by the distribution of absorbing material. This may be considered to be a first approximation. If the height of the tropopause is calculated from this approximation then it is in principle possible to calculate the temperature distribution in the stratosphere, and hence to evaluate the second term in equation (7).

T_r is clearly essentially positive, since the whole troposphere is at a higher temperature than the tropopause. According to (8), S_r is always negative. The magnitude of S_r will depend upon the ratio of the density of gas at the tropopause to the total quantity per cm.^{2} in the stratosphere. If there is a high density at the tropopause, and a small superincumbent blanket of gas to prevent the escape of radiation, then the cooling effect will be large. Such a state of affairs will normally be associated with a gas whose concentration falls off rapidly with height in the stratosphere. Therefore it may be expected that large negative values of S_r, i.e. large cooling effects, will be associated with the gases which are most rapidly attenuated in the stratosphere.

In order to visualize the way in which $\Sigma_r H_r$ varies with the height of the tropopause it is convenient to restrict the possible parameters as greatly as possible.

In the next section carbon dioxide is assumed to be thoroughly mixed with the atmosphere at all heights, which enables the density to be expressed in terms of atmospheric pressure and temperature. The concentration of water vapour in the troposphere can be conveniently related to air temperature by an assumption of constant percentage saturation, which accords roughly with observation. A direct relation between water-vapour concentration at the tropopause and total water vapour in the stratosphere is assumed which also accords roughly with such observations as exist. The concentration of ozone at the tropopause may be expressed in terms of latitude and season.

In principle this would allow $\Sigma_r H_r$ to be expressed in terms of latitude, season, ground temperature, stratosphere temperature and stratosphere pressure. Since this still represents a rather complicated system, these parameters are initially related as they occur in the earth's atmosphere, and the condition $\Sigma_r H_r = 0$ is found by successive approximations from the resulting figures. Thus figure 2 represents the energy unbalance in the earth's atmosphere as a function of latitude, but is used in a manner described to calculate the conditions which would lead to $\Sigma_r H_r = 0$ at all latitudes and seasons.

(b) *Data employed*

(i) Callendar (1941) found empirically a simple transmission function for regions of the 15μ carbon dioxide band, given by

$$\eta_r(\zeta) = \frac{1}{1+\bar{k}_r\zeta},$$

where ζ is the amount of absorbing material. This transmission function will be employed for all atmospheric gases, since it enables the expressions for T_r for water vapour and carbon dioxide to be integrated. Callendar allows for changing pressure by altering $\bar{k}_r$, and Elsasser follows a simpler procedure. It will therefore be assumed that the effect of a pressure gradient may be included by writing

$$\eta_r(x:y) = \left(1 + \int_x^y \bar{k}_r(z)\,\rho(z)\,dz\right)^{-1}.$$

$\bar{k}_r$ for water vapour and ozone are assumed to vary with pressure as $p^{\frac{1}{2}}$ in accordance with the theory of absorption by Lorentz lines. For carbon dioxide $\bar{k}_r$ is taken to vary as p. There is little evidence for or against such an assumption, and its effect is to produce a slight cooling effect by carbon dioxide in the tropics which would vanish if $\bar{k}_r$ were to vary as $p^{\frac{1}{2}}$.

(ii) Following Dobson and others, calculations will be restricted to ozone, carbon dioxide and water vapour. Oxides of nitrogen may have an effect similar to, but probably smaller than, ozone; lack of knowledge as to which oxides of nitrogen are present, and complete ignorance of their spatial distribution, makes it necessary to neglect their possible effect. A simple atmospheric absorption spectrum is shown in table 1 which will be used for computation.

Table 1

band centre (μ)	band width (μ)	gas	$\bar{k}_r$*	reference†
4·3	0·5	CO_2	1660	S.C.
4·75	0·5	O_3	2320	S.C.
6·3	1·8	H_2O	188	F.
9·65	0·4	O_3	2320	S.C.
14·1‡	3·0	O_3	155	S.C.
15	2·0	CO_2	755	S.C.
25	10	H_2O	80	H.
30 → ∞	—	H_2O	376	H.

* $\bar{k}_r$ for diffuse radiation $= 1{\cdot}66\bar{k}_r$ for parallel radiation (§ 2). The figures refer to laboratory conditions of temperature and pressure.

† S.C., Sutherland & Callendar (1942–3). These figures were evaluated using the Callendar transmission function. F., Fowle (1917). H., Hettner (1918). Results are corrected for temperature as suggested by Cowling (1942–3).

‡ Overlapped by 15μ CO_2. The effect of this band is halved in subsequent computations.

(iii) As a first approximation for calculation, and as a comparison with results obtained, a relation is required between ground temperature, stratosphere temperature and latitude as observed on the earth's surface. The southern hemisphere, although not so well observed as the northern hemisphere, shows ground temperatures and tropopause heights which are nearly independent of longitude. Ground temperatures given by Shaw (1928) and tropopause heights given by Dewar & Sawyer (1947) may be compounded with the relation $\partial\theta(t)/\partial x = -6{\cdot}5°\,\text{K/km.}$ to give the variation of $\theta(s)$ with latitude. The figures obtained correspond well with the figures given by Goldie (1947). Table 2 summarizes the figures adopted. With a steady lapse in the troposphere, and constant temperature in the stratosphere, the following relations hold between $p(x)$ (atmospheric pressure) and $\theta(x)$:

$$\frac{p(x)}{p(0)} = \left[\frac{\theta(x)}{\theta(0)}\right]^{5\cdot26} \quad \text{for the troposphere,}$$

$$\frac{p(x)}{p(s)} = \exp\left[-\frac{g}{R\theta(s)}(x-s)\right] \quad \text{for the stratosphere.}$$

(iv) The distribution of $\rho(x)\,\bar{k}_r(x)$ in the troposphere is required for calculation, rather than the distribution of $\rho(x)$. For the stratosphere only a knowledge of $\int_s^\infty \rho(x)\,\bar{k}_r(x)\,dx$ is required. Since the ozone content of the troposphere is small, it is only necessary to know $\rho(s)\,\bar{k}_r(s)$ to determine T_r for ozone (see § 4 (*c*) (ii), p. 498). Since nearly all ozone is concentrated in the stratosphere $\int_0^\infty \rho(x)\,\bar{k}_r(x)\,dx$ may be identified with $\int_s^\infty \rho(x)\,\bar{k}_r(x)\,dx$. Table 3 shows figures for $\rho(s)$, and $\int_0^\infty \rho(x)\,dx$ as a function of latitude and season. The figures for $\rho(s)$ are taken from Regener (1938), and the figures for $\int_0^\infty \rho(x)\,dx$ from Chapman (1942–3). The seasonal variation of $\rho(s)$ is inferred from Dobson, who believes that variations of density at the tropopause are twice as great as the variations of total ozone path. It is assumed that all figures vary linearly with latitude.

TABLE 2

latitude	tropopause height (km.)	ground temp. $\theta(0)$ (°K)	tropopause temp. $\theta(s)$ (°K)	tropopause pressure $p(s)$ (atm.)
0	17·8	302	186	0·079
10	16·5	300	193	0·098
20	14·8	296	200	0·127
30	12·9	291	207	0·167
40	11·2	284	211	0·210
50	9·7	278	215	0·260
60	8·5	272	218	0·307
70	7·6	267	219	0·339
80	7·0	264	220	0·375

TABLE 3

latitude	$\int_0^\infty \rho(x)\,dx$ g./cm.^{2} ($\times 10^{-4}$)	seasonal change of $\int_0^\infty \rho(x)\,dx$ (%)	$\rho(s)$ g./cm.^{3} ($\times 10^{-11}$)	seasonal change of $\rho(s)$ (%)
0	4·7	0	6	0
60	6·4	± 20	8·3	± 40

(v) Water vapour will be assumed to be 45 % saturated in the upper troposphere. This figure appears to be the best average available from measurements of frost-point temperature in the upper troposphere.

In § 4 (*c*) (iii) it is argued that substantial contributions to T_r for water vapour come from the uppermost few kilometres of the troposphere. In this region a frost-point lapse rate of 6·5° K/km. can be reasonably accurately represented by the density distribution

$$\bar{k}_r(x)\,\rho(x) = \bar{k}_r(s)\,\rho(s)\exp\left[\frac{6\cdot 65\times 10^3}{\theta(s)^2}\frac{\partial\theta(t)}{\partial x}(s-x)\right],$$

which includes the effect of $\bar{k}_r(x)$ varying as $p(x)^{\frac{1}{2}}$.

If the same distribution held for $x > s$, i.e. in the stratosphere, then

$$\frac{\int_s^\infty \bar{k}_r(x)\,\rho(x)\,dx}{\bar{k}_r(s)\,\rho(s)} = \frac{\theta(s)^2}{0\cdot 432} \approx 1\text{ km.}$$

Dobson supposes that in addition to an ice-point temperature lapse of 5 to 12° K/km. there is a minimum ice-point in the stratosphere of 190° K. For temperate latitudes 20 km. of such water-vapour concentration would correspond to 2 km. at the concentration which exists at the tropopause. Regener (1941) has made one humidity measurement at 22·2 km. and finds a relative humidity against ice of 7 % at a temperature of 226° K. This indicates a higher frost-point temperature than 190° K. As a rough estimate it will be assumed that

$$\int_s^\infty \bar{k}_r(x)\,\rho(x)\,dx = 4\bar{k}_r(s)\,\rho(s)\,\frac{\theta(s)^2}{0{\cdot}432},$$

i.e. the absorption is that of approximately 4 km. of air at the humidity of the tropopause.

(vi) Carbon dioxide will be assumed to form 0·03 % by volume of the atmosphere at all altitudes. This means $\rho(x) \propto p(x)/\theta(x)$.

With $\bar{k}_r(x) \propto p(x)$, then $\rho(x)\,\bar{k}_r(x) \propto \dfrac{p(x)^2}{\theta(x)}$.

If ground pressure is taken to be 1 atm. under all conditions, this gives all the information required about carbon dioxide.

(c) H_r *for the atmospheric gases*

(i) Substitution of the Callendar transmission function into equation (8) gives

$$S_r = -\bar{k}_r(s)\,\rho(s)\,B_r(s)\,\eta_r^2(s:\infty),$$

which expression will apply to all atmospheric gases.

(ii) For ozone the absorptive power of the troposphere is very small and

$$\frac{\partial \eta_r(x:s)}{\partial s} = -\rho(s)\,\bar{k}_r(s).$$

This allows T_r to be written in the simplified form

$$T_r = \bar{k}_r(s)\,\rho(s)\,\{B_r(0) - B_r(s)\}.$$

(iii) Using the Callendar transmission function and an exponential distribution of $\rho(x)\,\bar{k}_r(x)$, the expression (6) for T_r may be integrated, provided that $\partial B_r(x)/\partial x$ does not vary greatly from its value $\partial B_r(t)/\partial x$ at the tropopause. In temperate latitude gaseous concentrations are such that substantial contributions to T_r for water vapour come only from within the topmost 3 km. of the troposphere, and in this region the maximum variation of $\partial B_r(x)/\partial x$ is 3 % for the 30μ to ∞ region and 11 % for the 20 to 30μ region. No great error will therefore be involved by taking $\partial B_r(x)/\partial x = \partial B_r(t)/\partial x$, and also the contribution from the lower limit of the integral to T_r will be negligible. These simplifications enable T_r for water vapour to be written

$$T_r = \frac{\partial B_r(t)}{\partial x}\,\frac{\xi}{(1-\xi)^2}\,[1 - \xi + \log \xi], \tag{9}$$

where $\xi = \int_s^\infty \rho(x)\,\bar{k}_r(x)\,dx$ (not using the actual distribution in the stratosphere, but using the tropospheric distribution extrapolated to $x > s$).

(iv) Using the distribution appropriate to carbon dioxide does not lead to an integrable expression for T_r. If, however, $\theta(x)\dfrac{\partial B_r(x)}{\partial x}$ may be treated as constant in the troposphere and equal to $\theta(t)\dfrac{\partial B_r(t)}{\partial x}$, then T_r has precisely the same form (9) as for water vapour. In temperate latitudes the majority of the contribution to T_r comes from the last half kilometre of the troposphere, and in this region $\theta(x)\dfrac{\partial B_r(x)}{\partial x}$ for the 15μ band varies by less than 5 %. Thus for carbon dioxide as well as water vapour, the value of T_r will be taken from equation (9).

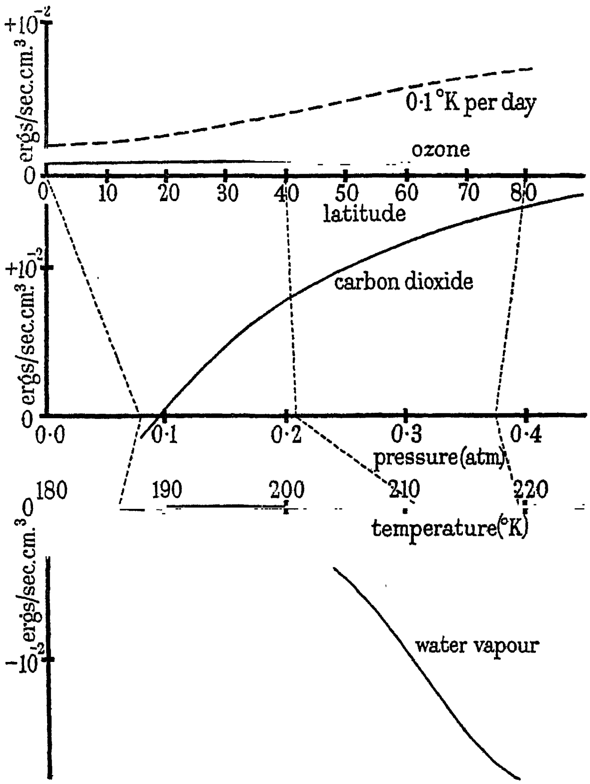


FIGURE 2. H_r for water vapour, carbon dioxide and ozone.

(v) Using the expressions and data given above H_r for the atmospheric gases may be calculated, and the results are shown in figure 2. By taking yearly average figures for ozone concentration and using the relations between $p(s)$, $\theta(s)$ and $\theta(0)$ given in table 2, the effect of ozone may be plotted as a function of latitude. H_{O_3} is almost exactly proportional to $\rho(s)$.

From the assumptions made, the effect of water vapour will be almost a function of tropopause temperature only. Table 2 was used to include the effect of pressure upon $\bar{k}_r(s)$, but a different relation between $\theta(s)$ and $p(s)$ would not noticeably affect the figures given. For this reason H_{H_2O} is plotted against $\theta(s)$. The effect of water vapour is almost entirely due to the rotation band, i.e. $\lambda = 20\mu$ to ∞.

The effect of carbon dioxide can only be described in terms of $p(s)$ and $\theta(s)$ together. The most important variable is, however, $p(s)$, which fully determines the value of ξ. The curve of H_{CO_2} against $p(s)$ is, therefore, dependent upon the $p(s) \sim \theta(s)$ relationship chosen. In constructing figure 2 the figures of table 2 are employed. The effect of carbon dioxide is almost entirely due to the 15μ band.

The dotted lines join values of $p(s)$, $\theta(s)$ and latitude which correspond in table 2. The dashed line shows the rate of gain of energy per cm.^{3} required to heat the lower stratosphere 0·1° K/day at various latitudes.

Figure 2 may be taken to represent the rate of gain or loss of energy in the lower stratosphere at various latitudes. Although there is not an exact balance at all latitudes, there are at least signs of how balance may be achieved. In temperate latitudes and in the arctic the balance will clearly be between heating by carbon dioxide and cooling by water vapour. In the tropics the heating may be due to ozone and the cooling due to carbon dioxide. More accurate data and more exact calculations may well show a balance to exist at all latitudes. In the next section figure 2 is treated as a first approximation in a calculation of tropopause height from the data available.

5. LATITUDE VARIATION OF STRATOSPHERE TEMPERATURE

Dobson has suggested that the difference of stratosphere temperatures in the tropics and the arctic may be accounted for by the higher ozone concentrations in the arctic. Figure 2, however, shows that despite the increase of zone concentration with latitude, and despite the increase of $\bar{k}_r(s)$ for ozone (owing to the higher tropopause pressures in the arctic), ozone supplies no greater heating effect in the arctic than in the tropics. This arises from the fact that the earth's surface is the heat source, and the decrease in surface temperature with latitude compensates the increase of $\rho(s)\,\bar{k}_r(s)$. The effect of ozone is small, and is probably important only in the tropics. This suggests the simplifying assumption that H_{O_3} is simply a constant under all circumstances if the total ozone content of the atmosphere is constant. In this case the position of the tropopause will be governed by

$$H_{H_2O}(\theta(s)) + H_{CO_2}(\theta(s), p(s)) = \text{const.},$$

which will lead to a relation between $p(s)$ and $\theta(s)$.

Despite possible inaccuracies in the calculations of § 4, the following statements are unlikely to be altered by improvement of data or calculating methods.

For the range of $p(s)$ and $\theta(s)$ observed in the earth's atmosphere:

(a) H_{H_2O} is a decreasing function of $\theta(s)$,

(b) H_{CO_2} is an increasing function of $\theta(s)$ and $p(s)$.

In figure 3 let $\theta_1(0)$, t_1, $\theta_1(s)$ be an equilibrium temperature configuration associated with a ground temperature $\theta_1(0)$. Consider the configuration resulting from a higher ground temperature $\theta_2(0)$. Assume that the new tropopause has reached t_2', so that the temperature distribution is given by $\theta_2(0)$, t_2', $\theta_1(s)$, and compare the balance of heating at t_1 and t_2'. The temperature is the same at t_1 and t_2', and the cooling due to water vapour must be the same at the two places. However, the pressure at t_2' will be lower than at t_1, and therefore the heating effect of carbon dioxide will be lower at t_2'. Since all heating and cooling contributions balance at t_1, there must be a net cooling at t_2'. Therefore the stratosphere at t_2' will start to cool, and the equilibrium value of $\theta_2(s)$ will be lower than $\theta_1(s)$. This simple argument illustrates the likelihood of a contraphase variation of $\theta(0)$ and $\theta(s)$ owing to the conflict of cooling effects by water vapour and heating effects by carbon dioxide.

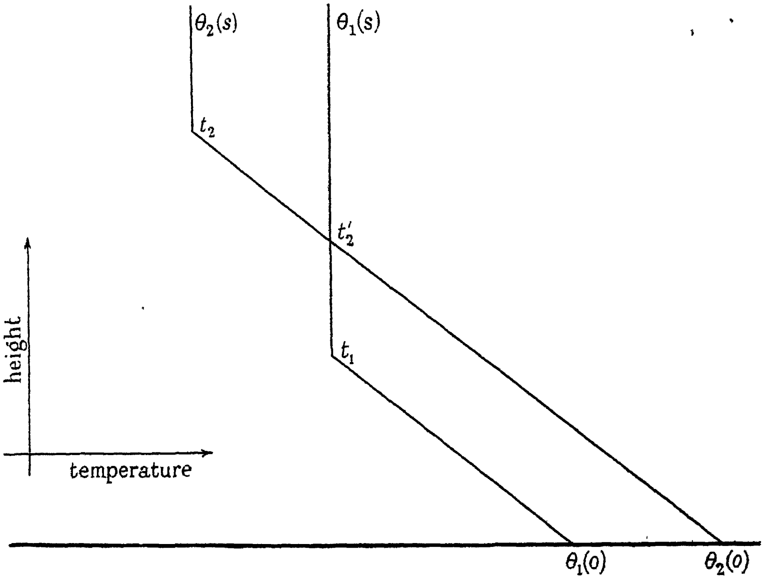


FIGURE 3. Latitude variation of $\theta(s)$.

It is possible to calculate from the figures of § 4, the exact $\theta(s)$, $p(s)$ relationship that will give $\Sigma_r H_r = 0$. This must be done by successive approximations since H_{CO_2} is a function of both variables. The results of such a calculation are shown in table 4.

TABLE 4

ground temperature (° K)	tropopause temperature (° K)		tropopause height (km.)	
	calculated	observed	calculated	observed
300	205	193	14·6	16·5
290	208	208	12·6	12·6
280	210	214	10·8	10·2
270	213	217	8·8	8·2
260	214	219	7·1	6·3

Table 4 shows only very rough agreement between calculated and observed values, but for $\theta(0) = 290°$ K the agreement is perfect, and the contraphase variation of $\theta(s)$ and $\theta(0)$ is clearly shown. Many possible causes for inaccuracy of the calculated values exist, but the most serious is probably the approximation of § 4 (*b*) (v) for the water-vapour content of the stratosphere. Variation of this approximation will not alter essentially the simple argument given above, but will alter the actual values calculated for $\theta(s)$.

Given the existence of a contraphase variation of $\theta(s)$ and $\theta(0)$, the simple argument given above with figure 3 may be used to show that the temperature configurations are stable, i.e. a displacement from an equilibrium state will alter the energy balance in such a way as to cause a return to the equilibrium state.

6. SEASONAL VARIATION OF STRATOSPHERE TEMPERATURE

Dobson has suggested that if the observed seasonal variation of ozone in temperate latitudes could cause an inphase variation of stratosphere temperature of ± 4° K, then the observed anomalous seasonal variation of stratosphere temperature would be explained. The theoretical variation of $\theta(s)$ with $\rho(s)_{\mathrm{O_3}}$ with $\theta(0)$ constant may be calculated from the data given in § 4. For simplicity, suppose $H_{\mathrm{O_3}}$ for $\theta(0)$ constant to be a function of $\rho(s)_{\mathrm{O_3}}$ only. A change of $\rho(s)_{\mathrm{O_3}}$ will cause changes of $p(s)$ and $\theta(s)$, and if thermal equilibrium is to be maintained the change must take place in such a way that

$$\Delta(H_{\mathrm{H_2O}}+H_{\mathrm{CO_2}}+H_{\mathrm{O_3}}) = 0.$$

Or, since $H_{\mathrm{O_3}} \propto \rho(s)_{\mathrm{O_3}}$,

$$-H_{\mathrm{O_3}}\frac{\Delta\rho(s)_{\mathrm{O_3}}}{\rho(s)_{\mathrm{O_3}}} = \Delta\theta(s)\left[\frac{\partial(H_{\mathrm{H_2O}}+H_{\mathrm{CO_2}})}{\partial\theta(s)}+\frac{\partial H_{\mathrm{CO_2}}}{\partial p(s)}\left(\frac{\partial p(s)}{\partial\theta(s)}\right)_{\theta(0)\,\mathrm{const.}}\right].$$

This expression contains the differential $\partial H_{\mathrm{H_2O}}/\partial\theta(s)$, which, as explained in the previous section, is the differential coefficient of probably the most uncertain quantity in these calculations. This quantity may be eliminated if the observed values of $\theta(s)$ and $p(s)$ in the earth's atmosphere are assumed. Consider a change of latitude denoted by Δ_l. If it is assumed, as in the previous section, that $H_{\mathrm{O_3}}$ does not vary, then

$$\Delta_l(H_{\mathrm{H_2O}}+H_{\mathrm{CO_2}}) = 0,$$

or

$$\frac{\partial(H_{\mathrm{H_2O}}+H_{\mathrm{CO_2}})}{\partial\theta(s)} = -\left(\frac{\partial p(s)}{\partial\theta(s)}\right)_l\frac{\partial H_{\mathrm{CO_2}}}{\partial p(s)}.$$

Therefore

$$H_{\mathrm{O_3}}\frac{\Delta\rho(s)_{\mathrm{O_3}}}{\rho(s)_{\mathrm{O_3}}} = \Delta\theta(s)\frac{\partial H_{\mathrm{CO_2}}}{\partial p(s)}\left[\left(\frac{\partial p(s)}{\partial\theta(s)}\right)_l-\left(\frac{\partial p(s)}{\partial\theta(s)}\right)_{\theta(0)\,\mathrm{const.}}\right].$$

$\left(\frac{\partial p(s)}{\partial\theta(s)}\right)_l$ is obtainable from table 2. $\left(\frac{\partial p(s)}{\partial\theta(s)}\right)_{\theta(0)\,\mathrm{const.}}$ will be $\frac{5\cdot26p(s)}{\theta(s)}$ if there is a constant lapse rate of temperature of 6·5° K/km. in the troposphere. The other quantities are all obtainable from the data given in § 4. The following values are found for

$$\theta(s) = 210^\circ\,\mathrm{K}, \quad p(s) = 0\cdot208\,\mathrm{atm.}:$$

$$\left(\frac{\partial p(s)}{\partial\theta(s)}\right)_l = 9\cdot2\times10^{-3}, \quad \left(\frac{\partial p(s)}{\partial\theta(s)}\right)_{\theta(0)\,\mathrm{const.}} = 5\cdot2\times10^{-3},$$

which give

$$\Delta\theta(s) = 4\cdot9\frac{\Delta\rho(s)_{\mathrm{O_3}}}{\rho(s)_{\mathrm{O_3}}} = \pm 2^\circ\,\mathrm{K}$$

for seasonal changes in temperate latitudes. Since all the quantities involved are rather uncertain, the agreement in order of magnitude with the observed changes gives encouragement to believe that Dobson's account of the seasonal variation of stratosphere temperature is correct.

7. TEMPERATURE GRADIENTS IN THE LOWER STRATOSPHERE

In the above sections the question of the distribution of water vapour in the stratosphere has been avoided owing to the serious uncertainties involved. In order to investigate the starting temperature gradient in the stratosphere some suitable assumption must be introduced. The calculation of $\partial\theta(s)/\partial x$ made below is only a second approximation, and owing to the uncertainty attached to the distribution of water vapour in the stratosphere, the figures given will do no more than illustrate the magnitude of the factors involved.

Using methods exactly similar to those employed in the derivation of (6) and (7), equation (1) at a general height y in the stratosphere may be put into the form

$$\Sigma_r\left\{\int_0^y \frac{\partial\eta_r(x:y)}{\partial y}\frac{\partial B_r(x)}{\partial x}\,dx - B_r(\infty)\frac{\partial\eta_r(y:\infty)}{\partial y} + \int_y^\infty \frac{\partial\eta_r(y:x)}{\partial y}\frac{\partial B_r(x)}{\partial x}\,dx\right\} = 0.$$

Differentiating with respect to y and setting $y = s$

$$-2\Sigma_r\,\bar{k}_r(s)\,\rho(s)\frac{\partial B_r(s)}{\partial x}$$

$$= \Sigma_r\left\{B_r(\infty)\frac{\partial^2\eta_r(s:\infty)}{\partial s^2} - \int_s^\infty \frac{\partial^2\eta_r(s:x)}{\partial s^2}\frac{\partial B_r(x)}{\partial x}\,dx - \int_0^s \frac{\partial^2\eta_r(x:s)}{\partial s^2}\frac{\partial B_r(x)}{\partial x}\,dx\right\}.$$

Using the Callendar transmission function the right-hand side may be written

$$= \Sigma_r\epsilon_r H_r + 2\Sigma_r\bar{k}_r(s)^2\rho(s)^2$$

$$\times\left\{B_r(\infty)\,\eta_r(s:\infty)^3 - \int_s^\infty \frac{\partial B_r(x)}{\partial x}\eta_r(s:x)^3\,dx - \int_0^s \frac{\partial B_r(x)}{\partial x}\eta_r(x:s)^3\,dx\right\},$$

where
$$\epsilon_r = -\frac{1}{\bar{k}_r(s)\,\rho(s)}\frac{\partial\bar{k}_r(s)\,\rho(s)}{\partial x}.$$

The left-hand side is directly proportional to $\partial\theta(s)/\partial x$, and this equation separates out the effects due to the upward flux of radiation and those due to the presence of more than one absorbing gas. The first term on the right-hand side would vanish if only one absorbing gas were present, for then there would only be one ϵ_r and the term would reduce to $\epsilon\Sigma_r H_r = 0$. The term would also vanish if all absorbing gases were similarly distributed, as would happen if they were mixed together by turbulence. It was pointed out in § 4 (*a*) that cooling effects (i.e. negative H_r) would be associated with the most rapidly attenuated gases (large positive ϵ_r). Thus since $\Sigma_r H_r = 0$ the weighting of the negative H_r by large ϵ_r will make $\Sigma_r\epsilon_r H_r$ negative. This factor alone would require a positive gradient of temperature in the lower stratosphere, and its contribution to $\dfrac{\partial\theta(s)}{\partial x}$ will be referred to as the 'distribution-term'.

The last three terms on the right-hand side are together equivalent to the term $k^2(s)\,\rho^2(s)\,\phi$ which would occur if the atmosphere were a grey absorber. Their contribution to $\dfrac{\partial\theta(s)}{\partial x}$ will therefore be referred to as the 'flux-term'. As is well known the flux term demands a negative gradient in the lower stratosphere, and the actual gradient may be regarded as the result of a conflict between the flux and distribution terms which differ in sign. The approximately isothermal character of the lower stratosphere indicates that the terms must either be small, or else of the same numerical magnitude.

Dobson has already in effect postulated a positive distribution term when he supposed ozone to be both the main heating agent at the tropopause, and the least rapidly attentuated absorbing gas. The figures of § 5 show that ozone is probably of little importance in temperate latitudes and in the arctic, but, nevertheless, it will be seen that cooling is due to the most rapidly attenuated gas (water vapour) and hence a positive distribution term may be expected at all latitudes. Indeed a positive distribution term may be expected in any atmosphere which has a convective and a radiative region.

The second and third terms on the right-hand side represent the contribution by the stratosphere to the flux term. Since this represents only a part of the flux term (in the case of carbon dioxide and water vapour the smaller part) it is reasonable to approximate to avoid the difficulties arising from the lack of knowledge of $\partial B_r(x)/\partial x$ in the stratosphere. A suitable approximation is to take from observation $\partial B_r(x)/\partial x = 0$ in the stratosphere, which causes the third term on the right-hand side to vanish. By the same approximation $B_r(\infty)$ must be replaced by $B_r(s)$. The effect of this approximation will be to overestimate slightly the numerical values of both the flux and the distribution term. Using this approximation, calculations may be made from the figures appropriate to the atmosphere calculated in § 5. The results of such a calculation are shown in table 5.

TABLE 5

	$\frac{\partial\theta(s)}{\partial x}$ (° K/km.)		
$\theta(s)$ (° K)	flux term	distribution term	sum
200	−3·0	+1·3	−1·7
210	−3·2	+1·1	−2·1
220	−3·3	+0·2	−3·1

Carbon dioxide is the main contributor to the flux term. Since the values of the tropopause pressure used correspond to those occurring in the arctic and temperate latitudes, it is probable that the figures shown in table 5 for the flux term are only comparable with the higher latitudes on the earth. In the tropics a considerably smaller value of the flux term would be expected, but rough calculations indicate that the distribution term would remain about 1 to 2° K/km. This at least indicates the probable cause of the small temperature gradients existing in the lower stratosphere.

The author is deeply indebted to Professor T. G. Cowling for the trouble he has taken in helping to prepare this paper for publication, to Dr G. B. B. M. Sutherland and Dr T. W. Wormell for encouragement and constructive criticism, and to the Ministry of Education for a Further Education and Training grant which rendered this work possible.

References

Callendar, G. S. 1941 *Quart. J. R. Met. Soc.* **67**, 263.
Chapman, S. 1942–3 *Rep. Progr. Phys.* **9**, 92.
Cowling, T. G. 1942–3 *Rep. Progr. Phys.* **9**, 29.
Dewar, D. & Sawyer, J. S. 1947 *M.R.P.* no. 334, Air Ministry, London.
Dobson, G. M. B. 1946 *Proc. Roy. Soc.* A, **185**, 144.
Elsasser, W. M. 1942 *Harv. Met. Stud.* no. 6.
Emden, R. 1913 *S.B. bayer. Akad. Wiss.* Heft 1.
Fowle, F. E. 1917 *Smithson Misc. Coll.* **68**, no. 8.
Gold, E. 1909 *Proc. Roy. Soc.* A, **82**, 43.
Goldie, A. H. R. 1947 *M.R.P.* no. 360, Air Ministry, London.
Hettner, G. 1918 *Ann. Phys., Lpz.*, (4), **55**, 476.
Hopf, E. 1934 *Mathematical problems of radiative equilibrium.* Cambridge University Press.
Paneth, F. A. & Glückauf, E. 1935 *Nature*, **136**, 717.
Regener, E. 1936 *Nature*, **138**, 544.
Regener, E. 1941 *Schr. dtsch. Akad. Luftfahrtforsch.* **46**, 5.
Regener, V. H. 1938 *Z. Phys.* **109**, 642.
Schwarzschild, D. K. 1906 *Nachr. Ges. Wiss. Göttingen*, **1**, 42.
Shaw, N. 1928 *Manual of Meteorology*, **2**. Cambridge University Press.
Sutherland, G. B. B. M. & Callendar, G. S. 1942–3 *Rep. Progr. Phys.* **9**, 18.

The scattering of high-energy neutrons by protons and the law of interaction between nucleons

By E. H. S. Burhop and H. N. Yadav, *University College, London*

(*Communicated by H. S. W. Massey, F.R.S.—Received* 10 *November* 1948)

Calculations have been carried out of the angular distribution and total collision cross-section to be expected in the case of the scattering of neutrons of energy in the range 83 to 250 MeV by protons. Born's approximation has been used in these calculations, and a large variety of possible shapes and ranges for the radial variation of the interaction potential have been investigated. For each case the radial variation of the central and non-central parts of the interaction was taken to be the same. The conditions for the validity of this approximation are discussed. Although only fair at the lower end of the energy range studied it should give reasonable results at the upper end of the range, especially in the case of potentials other than the spherical well.

The exchange properties of the interaction should be capable of deduction from the general shape of the angular distribution of the scattering curves. It is more difficult to obtain evidence about the exact form of the radial variation of the interaction potential from the angular distribution curves because, for any given potential shape, a wide variety of different predicted angular distributions are obtained by varying the range of the interaction. Potentials of the exponential and Yukawa type, however, show a greater concentration of the scattering into angles near 0 and 180° than is the case for a spherical well interaction.

The total collision cross-section depends quite markedly both on the exchange properties and radial variation of the interaction. Closest agreement with the measurements of total cross-section for 90 MeV neutrons are given for a symmetrical interaction with a radial variation in the form of a Yukawa potential of range $1{\cdot}18 \times 10^{-13}$ cm., although this interaction predicts too large a value for the ratio of the differential cross-section at 180° to that at 90°.

INTRODUCTION

Recent developments in cyclotron technique have made possible the production of beams of neutrons of energy of the order of 100 MeV or greater, and these are being used to investigate the total cross-section and angular distribution in the scattering of high-energy neutrons by protons. Such measurements are of fundamental importance because of the light they shed on the form of the law of interaction between nucleons. The law of interaction cannot, however, be obtained directly from such experiments. Instead, one has to calculate the cross-section and angular distributions expected on the assumption of a number of different forms of the interaction and to find which assumed interactions give results in closest agreement with experiment.

Camac & Bethe (1948) have shown that it is possible to explain the experimental results by assuming a spherical well shape of *n-p* interaction of range $2{\cdot}0 \times 10^{-13}$ cm. and by supposing the interaction to be a purely central one. Barker (1948) and Chew & Goldberger (1948) have also investigated the case of a purely central interaction, for potentials of the exponential and Yukawa form respectively.

If purely central forces only are considered however, difficulty is met with in trying to account for the observed electric quadrupole and magnetic moments of the deuteron. Rarita & Schwinger (1941) first considered the effect of non-central forces in the *n-p* scattering problem. The introduction of such forces not only adds greatly to the complexity of the calculation, but it also increases the ambiguity in the choice of a suitable form of the interaction energy. For example, should the radial dependence of the central and non-central interactions be the same?

It might be thought that meson theory would give a guide to the way in which the non-central interaction should be introduced, but in the various meson theories the non-central interaction is associated with singularities of a high order. Rarita & Schwinger assumed the radial dependence of both central and non-central interactions were the same, but they took three possible forms of the *n-p* interaction with exchange properties similar to those of the symmetrical, charged and neutral meson theories. They considered, however, only one form of the radial variation of the interaction energy—a spherical well of range $2{\cdot}80 \times 10^{-13}$ cm.

Recently several workers (Eisenstein & Rohrlich 1948; Ashkin & Wu 1948; Massey, Burhop & Hu 1948) have used the interaction constants deduced by Rarita & Schwinger to calculate the angular distribution and total collision cross-sections to be expected for collisions between neutrons of energy about 100 MeV with protons. Massey *et al.* (1948) have shown that with such a shape and range of interaction it is possible to get agreement with the early experimental results on the angular distribution, but the total cross-section for all types of interaction is too large.

A systematic investigation of the dependence of the total cross-section and angular distribution in high-energy *n-p* scattering on the shape and range of the assumed interaction has been made possible by the recent calculation by Hu & Massey (1948) of appropriate interaction constants for a large variety of such shapes and ranges.

Following Rarita & Schwinger, Hu & Massey assumed three types of interaction with exchange properties suggested by symmetrical, charged and neutral meson theories respectively.

They took the following three forms for the interaction energy:

$$\left.\begin{array}{l}\text{Interaction I (symmetrical theory):}\\ \mathscr{V} = (\hbar^2/Mr_0^2)\,\boldsymbol{\tau}_1\cdot\boldsymbol{\tau}_2[(a/12)\,(\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2+3)+(a\gamma S_{12}/3)+(g/4)\,(\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2-1)]\,V(r/r_0);\\ \text{Interaction II (charged theory):}\\ \mathscr{V} = (\hbar^2/Mr_0^2)\left(\dfrac{1+\boldsymbol{\tau}_1\cdot\boldsymbol{\tau}_2}{2}\right)[(a/4)\,(\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2+3)+a\gamma S_{12}+(g/4)\,(\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2-1)]\,V(r/r_0);\\ \text{Interaction III (neutral meson):}\\ \mathscr{V} = -(\hbar^2/Mr_0^2)\,[(a/4)\,(\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2+3)+a\gamma S_{12}-(g/4)\,(\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2-1)]\,V(r/r_0).\end{array}\right\}\quad(1)$$

In these expressions $\boldsymbol{\sigma}_1$, $\boldsymbol{\sigma}_2$, $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$ are the usual spin and isotopic spin operators; $S_{12} = \dfrac{3\boldsymbol{\sigma}_1\cdot\mathbf{r}\,\boldsymbol{\sigma}_2\cdot\mathbf{r}}{r^2} - \boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2$ is the tensor force operator; g, a, γ are dimensionless constants; and $V(r/r_0)$ a function that specifies the radial dependence of the interaction of 'range' r_0. In each case the form of $V(r/r_0)$ was taken to be the same for both the central and non-central parts of the interaction.

For even states all three interactions of (1) reduce to the same expressions, viz.

$$\left.\begin{array}{l}{}^1\mathscr{V}_{\text{even}} = -(\hbar^2/Mr_0^2)\,gV(r/r_0),\\ {}^3\mathscr{V}_{\text{even}} = -(\hbar^2/Mr_0^2)\,a(1+\gamma S_{12})\,V(r/r_0).\end{array}\right\}\quad(2)$$

Hu & Massey studied the following forms of the function $V(r/r_0)$ for a number of values of the range of interaction, r_0:

$$\left.\begin{array}{lll}\text{(1) Spherical well:} & V(r/r_0) = 1 & \text{for } r<r_0;\\ & = 0 & \text{for } r>r_0;\\ \text{(2) Yukawa potential:} & V(r/r_0) = (r_0/r)\,\mathrm{e}^{-r/r_0}; & \\ \text{(3) Exponential potential:} & & \\ & V(r/r_0) = \mathrm{e}^{-r/r_0}; & \\ \text{(4) Gaussian potential:} & V(r/r_0) = \mathrm{e}^{-(r/r_0)^2}. & \end{array}\right\}\quad(3)$$

In each case the constants a, g, γ were determined to give correctly the binding energy, the energy of the virtual singlet state, and the quadrupole moment of the deuteron.

Table 1 summarizes the cases studied and the interaction constants obtained as well as the similar constants for single cases derived from those of Rarita & Schwinger (1941) for the spherical well of range $2\cdot80\times10^{-13}$ cm. and of Chew & Goldberger (1948) for the Yukawa potential of range $1\cdot18\times10^{-13}$ cm. The constants shown in this table are those used in the present work.

The accurate calculation of the cross-sections and angular distributions in high-energy n-p scattering for these various forms of $V(r/r_0)$ is at present being carried out in this department but involves very lengthy computations. In the present work these cross-sections are calculated by means of Born's approximation for incident neutron energies in the range 83 to 250 MeV. The evidence available on the accuracy of the approximation under these conditions is discussed later. It would appear that it should give reasonable results except at the lower end of the energy range. At the upper end of this range one would expect relativistic effects to begin to assume importance.

TABLE 1

type of interaction	range r_0 ($\times 10^{-13}$ cm.)	a	g	$a\gamma$
spherical well	3·49	3·97	2·197	1·152
	3·05	3·24	2·231	1·620
	2·80	2·65	2·272	2·055
	2·62	1·93	2·262	2·49
	2·18	−1·83	2·295	4·78
Yukawa potential	2·18	2·43	1·487	1·069
	1·74	1·94	1·523	1·418
	1·18	0	1·575	2·879
exponential potential	1·09	8·2	5·176	3·935
	0·87	6·2	5·285	5·15
Gaussian potential	2·18	3·39	2·43	2·00

EXPRESSIONS FOR THE SCATTERING CROSS-SECTION

The expressions for the *n-p* scattering cross-section using Born's approximation and taking account of tensor forces has been derived by Ashkin & Wu (1948). Although our derivation differs a little from theirs it yields the same results, so that we discuss the theory only briefly in the present paper.

According to whether the neutron and proton spins are parallel or anti-parallel with respect to one another the states of the neutron-proton system will be triplets or singlets. If the symbols 1, 2 refer respectively to the spin co-ordinates of the neutron and proton the spin-dependent parts of the wave functions of the triplet states will be

$$\left.\begin{aligned} {}^3\chi_1 &= \alpha(1)\,\alpha(2), \\ {}^3\chi_0 &= \frac{1}{\sqrt{2}}\{\alpha(1)\,\beta(2)+\alpha(2)\,\beta(1)\}, \\ {}^3\chi_{-1} &= \beta(1)\,\beta(2), \end{aligned}\right\} \tag{4}$$

where the suffixes on the left-hand side refer to values of the magnetic quantum number m_s, and α, β are the ordinary eigenfunctions of the single-particle spin operator.

Similarly for the singlet case we have the spin-dependent part of the wave function given by

$${}^1\chi_0 = \frac{1}{\sqrt{2}}\{\alpha(1)\,\beta(2)-\alpha(2)\,\beta(1)\}. \tag{5}$$

Let $k\hbar\mathbf{n}_0$ be the initial momentum of the incident neutron in the centre of mass system and $k\hbar\mathbf{n}$ its momentum after scattering, $\mathbf{n}$ being unit vector specified by the co-latitude and azimuth θ, ϕ respectively, relative to $\mathbf{n}_0$. Since with the non-central interaction the total spin of the system remains a constant of the motion, the multiplicity of the state will not change in the scattering process, but the magnetic quantum number m_s may change.

Born's approximation gives for the differential cross-section for scattering into the solid angle $d\omega$ in the direction $\mathbf{n}(\theta,\phi)$ from an initial state specified by magnetic quantum number m_s to a final state of magnetic quantum number m_s',

$$I_{m_s' m_s}(\theta,\phi)\,d\omega = \left|\frac{M}{4\pi\hbar^2}\Sigma\int e^{ik(\mathbf{n}_0-\mathbf{n})\cdot\mathbf{r}}\chi_{m_s'}\mathscr{V}\chi_{m_s}\,d\tau\right|^2 d\omega, \tag{6}$$

the integration being carried out over all values of the spatial co-ordinates, the summation over all values of the spin co-ordinates.

TABLE 2. VALUES OF $\mathscr{V}_{m_s m_s'}$ FOR DIFFERENT VALUES OF m_s, m_s'

$m_s \backslash m_s'$	1	0	−1
1	$a\{1+2\gamma P_2(\cos\xi)\}$	$\sqrt{2}\,a\gamma P_2^1(\cos\xi)\,e^{-i\eta}$	$a\gamma P_2^2(\cos\xi)\,e^{-2i\eta}$
0	$\sqrt{2}\,a\gamma P_2^1(\cos\xi)\,e^{i\eta}$	$a\{1-4\gamma P_2(\cos\xi)\}$	$-\sqrt{2}\,a\gamma P_2^1(\cos\xi)\,e^{-i\eta}$
−1	$a\gamma P_2^2(\cos\xi)\,e^{2i\eta}$	$-\sqrt{2}\,a\gamma P_2^1(\cos\xi)\,e^{i\eta}$	$a\{1+2\gamma P_2(\cos\xi)\}$

For an initial state of quantum number m_s the complete differential cross-section is obtained by summing over all possible values of m_s' for the different final states. For a beam of polarized neutrons an average has to be taken over all initial states, so that we have for the triplet and singlet cases respectively,

$$I_{\text{trip.}}(\theta,\phi) = \frac{1}{3}\sum_{m_s=1,0,-1}\ \sum_{m_s'=1,0,-1} I^{\text{trip.}}_{m_s' m_s}(\theta,\phi), \tag{7}$$

$$I_{\text{sing.}}(\theta,\phi) = I^{\text{sing.}}_{00}(\theta,\phi). \tag{8}$$

Finally, for an actual beam of unpolarized neutrons the weighted mean of (7) and (8) has to be taken, giving for the expected differential cross-section

$$I(\theta,\phi) = \tfrac{3}{4}I_{\text{trip.}}(\theta,\phi) + \tfrac{1}{4}I_{\text{sing.}}(\theta,\phi). \tag{9}$$

To evaluate (6) it is necessary to calculate

$$\frac{M}{4\pi\hbar^2}\sum_{\substack{\text{spin}\\\text{co-ord.}}}\chi^*_{m_s'}\mathscr{V}\chi_{m_s}$$

for the three types of interaction (1).

Consider first the case of an interaction consisting entirely of ordinary forces (type III interaction). Writing

$$\frac{M}{4\pi\hbar^2}\sum_{\substack{\text{spin}\\\text{co-ord.}}}\chi^*_{m_s'}\mathscr{V}\chi_{m_s} = -\frac{1}{4\pi r_0^2}V(r/r_0)\,\mathscr{V}_{m_s m_s'}, \tag{10}$$

table 2 gives $\mathscr{V}_{m_s m_s'}$ for the triplet interaction for various values of m_s, m_s'. The spherical polar co-ordinates of the vector $\mathbf{r}$ have been taken as (r, ξ, η). For the

TABLE 3

	$c(x)/r_0$				$f(x)/r_0$			
x	spherical well	exponential well	Yukawa potential	Gaussian potential	spherical well	exponential well	Yukawa potential	Gaussian potential
0	0	0	0	0	0·33333	2·00000	1·00000	0·44310
0·1	—	0·01566	0·00394	0·00096	—	1·96061	0·99010	0·44200
0·2	0·00053	0·05886	0·01510	0·00190	0·33200	1·84911	0·96154	0·43869
0·3	—	0·12012	0·03186	0·00404	—	1·68336	0·91743	0·43324
0·4	0·00212	0·18582	0·05197	0·00699	0·32803	1·48633	0·86207	0·42573
0·5	—	0·24640	0·07360	0·01062	—	1·28000	0·80000	0·41626
0·6	0·00471	0·29680	0·09248	0·01497	0·32130	1·08131	0·73529	0·40497
0·7	—	0·33151	0·10991	0·01992	—	0·90086	0·67114	0·39201
0·8	0·00826	0·35149	0·12442	0·02531	0·31250	0·74360	0·60976	0·37759
0·9	—	0·35891	0·13557	0·03110	—	0·61048	0·55249	0·36188
1·0	0·01267	0·35620	0·14380	0·03712	0·30120	0·50000	0·50000	0·34509
1·1	—	0·34618	0·14933	0·04328	—	0·40949	0·45249	0·32744
1·2	0·01788	0·33127	0·15248	0·04949	0·28767	0·33591	0·40984	0·30914
1·3	—	0·31327	0·15383	0·05482	—	0·27639	0·37175	0·29041
1·4	0·02369	0·29381	0·15359	0·06155	0·27238	0·22827	0·33784	0·27145
1·5	—	0·27399	0·15204	0·06721	—	0·18935	0·30769	0·25247
1·6	0·02991	0·25437	0·14962	0·07250	0·25557	0·15781	0·28090	0·23365
1·7	—	0·23548	0·14649	0·07738	—	0·13217	0·25707	0·21514
1·8	0·03652	0·21754	0·14291	0·08175	0·23711	0·11125	0·23585	0·19712
1·9	—	0·20075	0·13897	0·08559	—	0·09411	0·21692	0·17970
2·0	0·04335	0·18516	0·13484	0·08885	0·21769	0·08000	0·20000	0·16301
2·1	—	0·17082	0·13054	0·09153	—	0·06833	0·18484	0·14713
2·2	0·05016	0·15759	0·12623	0·09360	0·19752	0·05864	0·17123	0·13213
2·3	—	0·14548	0·12194	0·09507	—	0·05055	0·15898	0·11807
2·4	0·05684	0·13439	0·11769	0·09597	0·17689	0·04377	0·14793	0·10498
2·5	—	0·12428	0·11353	0·09629	—	0·03805	0·13793	0·09288
2·6	0·06323	0·11504	0·10948	0·09608	0·15609	0·03321	0·12887	0·08176
2·7	—	0·10661	0·10554	0·09538	—	0·02910	0·12063	0·07161
2·8	0·06916	0·09891	0·10174	0·09424	0·13544	0·02559	0·11312	0·06241
2·9	—	0·09187	0·09808	0·09269	—	0·02259	0·10627	0·05412
3·0	0·07451	0·08545	0·09455	0·09077	0·11523	0·02000	0·10000	0·04670
3·1	—	0·07957	0·09116	0·08856	—	0·01777	0·09425	0·04010

3·2	0·07914	0·07419	0·08792	0·08608	0·09571	0·01583	0·08897	0·03425
3·3	—	0·06925	0·08481	0·08340	—	0·01415	0·08410	0·02912
3·4	0·08296	0·06472	0·08184	0·08084	0·07713	0·01268	0·07962	0·02463
3·5	—	0·06057	0·07899	0·07757	—	0·01139	0·07547	0·02072
3·6	0·08589	0·05673	0·07163	0·07451	0·05971	0·01026	0·07163	0·01735
3·7	—	—	0·06807	0·07141	—	—	0·06807	0·01446
3·8	0·08788	0·04997	0·06477	0·06829	0·04363	0·00839	0·06477	0·01198
3·9	—	—	0·06169	0·06519	—	—	0·06169	0·00989
4·0	0·08887	0·04420	0·06653	0·06212	0·02903	0·00692	0·05882	0·00812
4·2	0·08884	0·03926	0·06228	0·05619	0·01603	0·00575	0·05365	0·00539
4·4	0·08783	0·03502	0·05839	0·05063	0·00470	0·00482	0·04912	0·00350
4·6	0·08585	0·03134	0·05484	0·04548	−0·00491	0·00407	0·04513	0·00223
4·8	0·08296	0·02816	0·05157	0·04086	−0·01281	0·00346	0·04160	0·00140
5·0	0·07923	0·02539	0·04858	0·03662	−0·01902	0·00296	0·03846	0·00086
5·2	0·07476	0·02296	0·04583	0·03288	−0·02361	0·00254	0·03566	0·00051
5·4	0·06964	0·02083	0·04328	0·02956	−0·02667	0·00220	0·03316	0·00030
5·6	0·06400	0·01895	0·04094	0·02662	−0·02833	0·00191	0·03090	0·00017
5·8	0·05795	0·01729	0·03878	0·02404	−0·02870	0·00167	0·02887	0·00010
6·0	0·05163	0·01581	0·03678	0·02176	−0·02797	0·00146	0·02703	0·00005
6·2	0·04518	0·01450	0·03493	0·01974	−0·02627	0·00129	0·02535	0·00003
6·4	0·03871	0·01332	0·03321	0·01796	−0·02380	0·00114	0·02383	0·00002
6·6	0·03236	0·01227	0·03161	0·01638	−0·02073	0·00101	0·02244	0·00001
6·8	0·02624	0·01132	0·03012	0·01499	−0·01723	0·00090	0·02117	—
7·0	0·02045	0·01047	0·02872	0·01374	−0·01347	0·00080	0·02000	—
7·2	0·01509	0·00971	0·02742	0·01262	−0·00961	0·00072	0·01892	—
7·4	0·01022	0·00901	0·02621	0·01163	−0·00579	0·00064	0·01792	—
7·6	0·00594	0·00838	0·02508	0·01074	−0·00215	0·00058	0·01699	—
7·8	0·00226	0·00780	—	0·00993	+0·00122	0·00052	—	—
8·0	−0·00078	—	—	0·00920	0·00421	—	—	—
8·2	−0·00317	—	—	—	0·00675	—	—	—
8·4	−0·00493	—	—	—	0·00880	—	—	—
8·6	−0·00607	—	—	—	0·01033	—	—	—
8·8	−0·00663	—	—	—	0·01133	—	—	—
9·0	−0·00666	—	—	—	0·01181	—	—	—
9·2	−0·00622	—	—	—	0·01180	—	—	—
9·4	−0·00538	—	—	—	0·01134	—	—	—
9·6	−0·00422	—	—	—	0·01049	—	—	—
9·8	−0·00284	—	—	—	0·00930	—	—	—
10·0	−0·00124	—	—	—	0·00785	—	—	—

singlet case, $m_s = m_s' = 0$ and $\mathscr{V}_{00} = g$. To evaluate (6) we then need to know the integrals

$$\left.\begin{aligned}
&\int e^{ik(\mathbf{n}_0-\mathbf{n}).\mathbf{r}}\, V(r/r_0)\, d\tau && (= 4\pi r_0^2 f(\theta)),\\
&\int e^{ik(\mathbf{n}_0-\mathbf{n}).\mathbf{r}}\, V(r/r_0)\, P_2(\cos\xi)\, d\tau && (= 4\pi r_0^2\, c(\theta)),\\
&\int e^{ik(\mathbf{n}_0-\mathbf{n}).\mathbf{r}}\, V(r/r_0)\, P_2^1(\cos\xi)\, d\tau && (= 4\pi r_0^2\, c^{(1)}(\theta,\phi)),\\
&\int e^{ik(\mathbf{n}_0-\mathbf{n}).\mathbf{r}}\, V(r/r_0)\, P_2^2(\cos\xi)\, d\tau && (= 4\pi r_0^2\, c^{(2)}(\theta,\phi)).
\end{aligned}\right\} \tag{11}$$

Carrying out the integrations of (11) we obtain

$$r_0^2 f(\theta) = \int_0^\infty \frac{\sin \mathscr{K} r}{\mathscr{K} r}\, V(r/r_0)\, r^2\, dr, \tag{12}$$

$$r_0^2\, c(\theta) = P_2(\cos\vartheta) \int_0^\infty \left\{\frac{3(\sin \mathscr{K} r - \mathscr{K} r \cos \mathscr{K} r)}{\mathscr{K}^3 r^3} - \frac{\sin \mathscr{K} r}{\mathscr{K} r}\right\} V(r/r_0)\, r^2\, dr. \tag{13}$$

$c^{(1)}(\theta,\phi)$, $c^{(2)}(\theta,\phi)$ are similar to $c(\theta)$ but with $P_2(\cos\vartheta)$ replaced by $P_2^1(\cos\vartheta)\, e^{i\phi'}$, $P_2^2(\cos\vartheta)\, e^{2i\phi'}$ respectively. In these expressions $\mathscr{K} = 2k \sin\frac{1}{2}\theta$ and $2\vartheta = \pi - \theta$. ϕ' is an azimuthal co-ordinate of the vector $\mathbf{n}_0$ with respect to a polar axis in the direction of $\mathbf{n}_0 - \mathbf{n}$.

Carrying out the averaging process of equations (7) and (8) we obtain

$$I_{\text{trip.}}(\theta,\phi) = a^2\{f^2(\theta) + 8\gamma^2 c^2(\theta)\}, \tag{14}$$

$$I_{\text{sing.}}(\theta,\phi) = g^2 f^2(\theta), \tag{15}$$

so that the weighted mean of these gives for $I(\theta,\phi)$,

$$I(\theta,\phi) = \tfrac{1}{4}(3a^2 + g^2) f^2(\theta) + 6\gamma^2 a^2 c^2(\theta), \tag{16}$$

which is the same as the expression obtained by Ashkin & Wu (1948).

For exchange forces of type II, Born's approximation gives an expression identical with (16) but with θ replaced by $\pi - \theta$, viz.

$$I(\theta,\phi) = \tfrac{1}{4}(3a^2 + g^2) f^2(\pi - \theta) + 6\gamma^2 a^2 c^2(\pi - \theta). \tag{17}$$

The total collision cross-section in this case is then the same as in the type III interaction, the differential cross-section at a scattering angle θ in the type II interaction being identical with that at $\pi - \theta$ in the type III interaction. This relation between the angular distributions in the type II and type III interaction is a feature of Born's approximation. It does not hold in the case of the exact calculation.

For the case of symmetrical forces of type I interaction, in the integrals of equation (11), e^{ikz} has to be replaced by $\cos kz - 3i \sin kz$ in the case of singlet states and by $\cos kz - \frac{1}{3}i \sin kz$ in the case of triplet states, where $z = \mathbf{n}_0 . \mathbf{r}$. This leads to the following expression for $I(\theta,\phi)$ for type I forces:

$$\begin{aligned}
I(\theta,\phi) = (g^2/4)\{f(\theta) - 2f(\pi - \theta)\}^2 &+ (a^2/12)\{f(\theta) + 2f(\pi - \theta)\}^2\\
&+ (2a^2\gamma^2/3)\{c^2(\theta) + 4c^2(\pi - \theta) - 2c(\theta)\, c(\pi - \theta)\}.
\end{aligned} \tag{18}$$

As pointed out by Ashkin & Wu for a given shape of interaction, $I(\theta,\phi)$ can be determined for any incident momentum $k\hbar$ and any angle of scattering θ when the functions $f(\theta)/r_0$, $c(\theta)/r_0$ are known because these are dimensionless quantities which involve θ through the product $\mathscr{K}r_0$ only. For the interaction shapes given in equation (3) the following expressions have been obtained for $f(\theta)$, $c(\theta)$ in terms of tabulated functions.

(1) Spherical well:

$$\left.\begin{aligned} f(\theta) &= (r_0/x)(\sin x - x\cos x), \quad x = \mathscr{K}r_0, \\ c(\theta) &= (r_0/x^3)(3\sin x - 4\sin x + x\cos x). \end{aligned}\right\} \tag{19}$$

These expressions agree with those given by Ashkin & Wu.

(2) Yukawa potential:

$$\left.\begin{aligned} f(\theta) &= r_0(1+x_0^2)^{-1}, \\ c(\theta) &= r_0\{3/x^2 - (3/x^3)\arctan x - (1+x^2)^{-1}\}. \end{aligned}\right\} \tag{20}$$

(3) Exponential well:

$$\left.\begin{aligned} f(\theta) &= 2r_0(1+x^2)^{-2}, \\ c(\theta) &= r_0\{(3/x^3)\arctan x - (3/x^2)(1+x^2)^{-1} - 2(1+x^2)^{-2}\}. \end{aligned}\right\} \tag{21}$$

(4) Gaussian potential:

$$\left.\begin{aligned} f(\theta) &= (r_0\pi^{\frac{1}{2}}/4)\,\mathrm{e}^{-x^2/4}, \\ c(\theta) &= (r_0\pi/8)\{(12/x^3)\,\mathrm{erf}\,(x/2) - 2(1+6/x^2)\,\pi^{-\frac{1}{2}}\,\mathrm{e}^{-x^2/4}\}. \end{aligned}\right\} \tag{22}$$

Table 3 gives the functions $f(\theta)/r_0$, $c(\theta)/r_0$ given by equations (19) to (22) for the four kinds of interaction considered. In this table $x = 2kr_0\sin\frac{1}{2}\theta$ is taken as argument. Knowing these functions the differential cross-section for any value of θ may be calculated for the three types of interaction and for any k, r_0, or shape of well, using equations (16) to (18).

VALIDITY OF BORN'S APPROXIMATION

Figure 1 shows a comparison between the calculated scattering curves for neutrons of energy 83 MeV calculated by means of Born's approximation and the accurate calculations of Massey *et al.* (1948) for a spherical well interaction of range $2{\cdot}80\times10^{-13}$ cm., while table 4 shows a similar comparison of total cross-sections. The agreement is only fair, being poorest in the case of an interaction of type II.

TABLE 4. TOTAL CROSS-SECTION ($\times 10^{-26}$ cm.^{2})

	interaction I	interaction II	interaction III
exact theory	14·10	15·77	23·38
Born approximation	19·05	23·32	23·32

The calculations of Ashkin & Wu show, however, that the position is quite different for neutrons of energy 200 MeV and the same shape of potential well. In that case the total cross-section given by Born's approximation for type I interaction agrees with the accurate cross-section to within 1 %, and the angular distributions are reasonably close.

Born's approximation might be expected to give rather better results for the other shapes of interaction. Interactions of other shapes remain finite out to greater distances than in the case of the spherical well, so that there is a greater contribution to the scattering from more distant collisions in these cases. As a result phase shifts, δ_l, corresponding to larger values of the angular momentum $l\hbar$, are important in calculating the scattering. But the larger the value of l, the more closely do the accurately calculated phase shifts agree with those of Born's approximation. So that the approximation might be expected to be the more accurate the larger the number of phases making a significant contribution to the scattering.* Applying this argument to the shapes of interaction discussed in this paper, one might expect the approximation to give better results for the Yukawa potential than for the Gaussian or spherical well interaction.

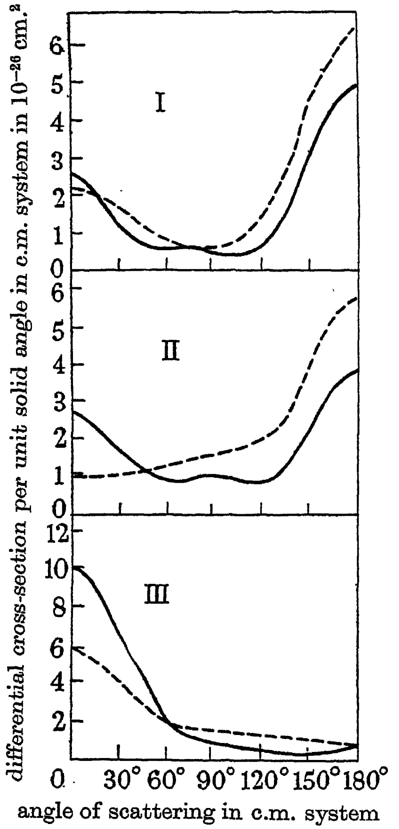


Figure 1. Comparison of angular distribution of scattered neutrons using exact calculation (full lines) and Born's approximation (broken lines) for neutrons of energy 83 MeV, assuming spherical well interaction of range $2{\cdot}80 \times 10^{-13}$ cm. Interactions, I, II and III.

Using a similar argument, for a given shape of interaction, Born's approximation might be expected to give more accurate results the longer the range.

---

* This argument assumes that the phase for small l are not less accurate in the Yukawa well case.

RESULTS OF THE CALCULATION

(a) *The functions $f(\theta)$, $c(\theta)$*

The form of the angular distribution of the scattered particles is determined by the form of the dimensionless quantities $f(\theta)/r_0$, $c(\theta)/r_0$, tabulated in table 3. The quantity $f(\theta)/r_0$ shows a more rapid decrease with increasing x for small x in the case of potential shapes (2) and (3) than in the case of potential shapes (1) and (4). This gives rise to a tendency to concentrate the scattering into small angles for these shapes of potential in the case of ordinary forces, and into angles near 180° in the case of exchange forces. Physically this concentration of the scattering into small angles in the case of the Yukawa and exponential well potentials arises from the long 'tail' of the potential function in these cases in comparison with the spherical well and Gaussian potential types of interaction.

For the spherical well type of potential both $f(\theta)$ and $c(\theta)$ pass through zero and become negative. The other types of potential do not show this behaviour. As a result the angular distribution is much less smooth for the spherical well than in the case of the other kinds of potentials.

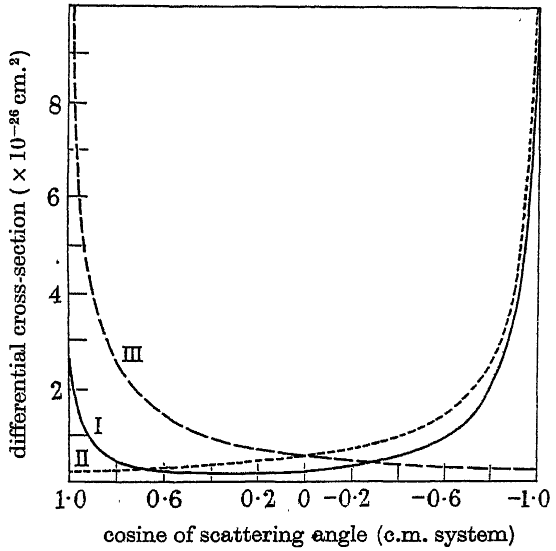


FIGURE 2. Angular distribution of scattered neutrons for three interaction types for Yukawa well range $1{\cdot}74\times10^{-13}$ cm. and incident neutron energy 100 MeV. —— interaction I, ----- interaction II, – – – – interaction III.

(b) *The angular distribution of the scattered neutrons*

Table 5 shows the calculated angular distribution in the scattering of neutrons of energies 83, 150, 250 MeV for different assumptions about the type of interaction and the shape and range of the potential.

The shapes of the angular distribution are qualitatively different for the three types of interaction assumed. This is illustrated in figure 2 which shows the angular distribution of scattered neutrons for the three interaction types for a Yukawa interaction of range $1{\cdot}74\times10^{-13}$ cm., and incident neutron energy 100 MeV. Thus a study of the angular distribution can enable a choice to be made between the different interaction types. The results of the preliminary experiments already reported using neutrons of energy 90 ± 13 MeV enable one to reject an interaction of type III, although more careful measurements and more accurate calculations will be required before a decision can be arrived at as to the actual exchange properties of the interaction.

Figure 3 shows the angular distribution for neutron energies of 83, 150 and 250 MeV for a Yukawa interaction of range $1\cdot18\times10^{-13}$ cm. and type I interaction. As the neutron energy increases the ratio of the scattering at 90° to the scattering at 180° decreases markedly. This is a characteristic feature for all shapes of well as is seen from table 5.

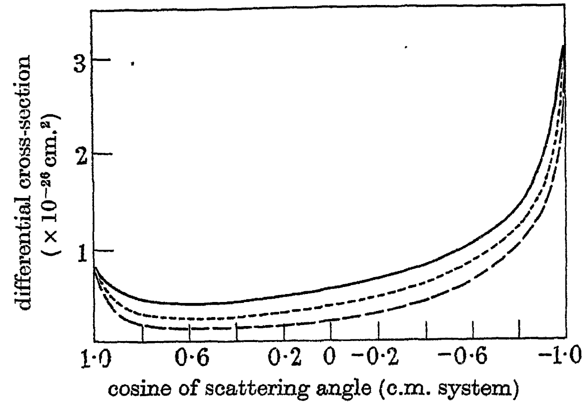


FIGURE 3. Angular distribution of scattered neutrons for three different neutron energies for Yukawa well range $1\cdot18\times10^{-13}$ cm. and type I interaction. Neutron energies: —— 83 MeV, -·-·- 150 MeV, ---- 250 MeV.

It is more difficult to get information regarding the shape of the potential from the angular distribution. For each shape of interaction there is a wide variety of forms of the angular distribution predicted, corresponding to different assumed ranges of interaction. This is shown particularly by figure 4 which shows, for incident neutrons of energy 83 MeV and symmetrical forces the shapes of calculated angular distributions assuming three different ranges for a spherical well potential. As the range is decreased the amount of non-central interaction necessary to give the correct deuteron quadrupole moment becomes greater. The effect of this extra non-central interaction is to give rise to more scattering near 90° in the centre of mass system.

If one considers different shapes of interaction and chooses a suitable range for each so as to make the ratio of the differential cross-section at 90° to that at 180° the same in each case, there are certain differences in the detailed form of the angular distributions that might be used to distinguish between the different interaction shapes. There is a tendency to a greater concentration of the scattering into angles near 0 and 180° for the Yukawa and exponential potentials than is the case for spherical well and Gaussian shapes of interaction. This is shown, for example, in figure 5, which illustrates the angular distribution for a spherical well interaction of range $2{\cdot}62 \times 10^{-13}$ cm. and a Yukawa potential of range $1{\cdot}18 \times 10^{-13}$ cm. The scale for the ordinates of these curves is arbitrarily adjusted to make the cross-section at 180° coincide in the two cases. When this is done the curves lie fairly close together at 90°, but the detailed shapes of the curves are quite different. Careful measurement of the angular distribution near 0 and 180° should enable a decision between these two possible potential shapes to be arrived at.

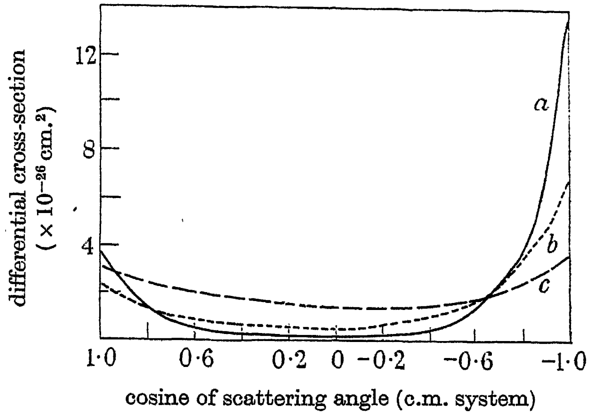


FIGURE 4. Angular distribution for spherical well shape of interaction and three different ranges. Neutron energy 83 MeV. Type I interaction. *a*, $3{\cdot}49 \times 10^{-13}$ cm.; *b*, $2{\cdot}80 \times 10^{-13}$ cm.; *c*, $2{\cdot}18 \times 10^{-13}$ cm.

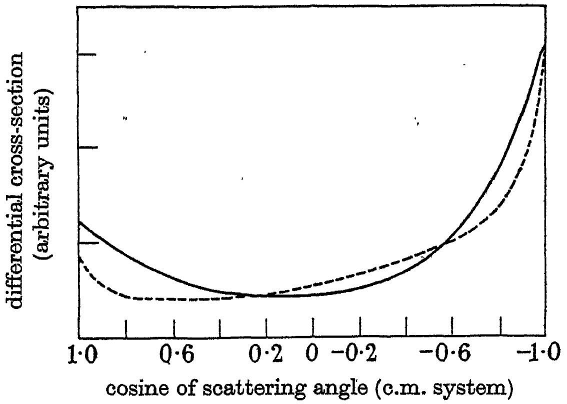


FIGURE 5. Shape of angular distribution for two potential shapes that give nearly the same ratio of scattering at 180° to scattering at 90°. The two curves are adjusted to fit arbitrarily at 180°. Neutron energy 83 MeV. Type I interaction: —— spherical well, range $2{\cdot}62 \times 10^{-13}$ cm; ----- Yukawa potential, range $1{\cdot}18 \times 10^{-13}$ cm.

As already pointed out, some of the angular distributions in the case of the spherical well potential show characteristic irregularities. These are more evident in the case of type III interaction, but they are also present in some of the type I interaction cases as can be seen from table 5, and could perhaps provide a means of distinguishing a potential shape of this type from other possibilities.

Table 5

shape of well	range of inter-action (× 10^{-13} cm.)	type of inter-action	neutron energy in laboratory system (MeV)	differential cross-section for scattering through angle θ (× 10^{-26} cm.^{2}) cos θ (θ in c.m. system) 1·0	0·95	0·9	0·8	0·6	0·4	0	−0·4	−0·6	−0·8	−0·9	−0·95	−1·0
spherical well	3·49	I	83	3·48	—	—	1·30	0·50	0·28	0·25	0·51	1·55	4·88	—	—	13·68
	3·05	I	83	2·53	2·14	1·83	1·32	0·71	0·48	0·39	0·90	1·80	4·17	6·14	7·41	8·96
	2·80	I	83	2·25	1·98	1·75	1·36	0·74	0·69	0·57	1·11	1·52	3·59	4·76	5·52	6·42
	2·62	I	83	2·07	1·88	1·71	1·43	1·04	0·83	0·72	1·14	1·78	3·01	3·96	4·55	5·23
	2·18	I	83	3·04	2·88	2·75	2·50	2·11	1·82	1·48	1·55	1·89	2·60	3·16	3·49	3·89
Yukawa	2·18	I	83	5·01	2·34	1·40	0·70	0·35	0·26	0·31	0·61	1·09	2·64	5·46	9·24	19·83
	1·74	I	83	2·64	1·55	1·05	0·61	0·35	0·30	0·37	0·70	1·15	2·40	4·25	6·30	10·67
	1·18	I	83	0·85	0·66	0·54	0·44	0·40	0·43	0·56	0·80	1·00	1·40	1·86	2·30	3·05
exponential	1·09	I	83	3·69	2·49	1·75	1·02	0·35	0·23	0·27	0·64	1·27	3·30	6·35	9·31	14·60
	0·87	I	83	1·95	1·45	1·12	0·74	0·43	0·35	0·44	0·85	1·38	2·74	4·29	5·58	7·54
Gaussian	2·18	I	83	2·39	1·90	1·52	0·99	0·50	0·34	0·38	0·87	1·64	3·63	5·69	7·18	9·10
spherical well	3·49	III	83	17·63	—	—	6·39	2·27	1·03	0·67	0·44	0·29	0·16	—	—	0·07
	3·05	III	83	9·42	7·82	6·49	4·52	2·30	1·55	1·15	0·90	0·74	0·57	0·50	0·45	0·41
	2·80	III	83	5·71	4·88	4·21	3·20	1·92	1·75	1·61	1·43	1·26	1·06	0·95	0·90	0·85
	2·62	III	83	3·11	2·72	2·42	2·01	1·71	1·75	2·00	1·98	1·83	1·64	1·52	1·47	1·41
	2·18	III	83	2·02	1·86	1·76	1·71	2·03	2·65	4·01	4·91	5·10	5·14	5·11	5·08	5·04
Yukawa	2·18	III	83	23·67	11·26	6·89	3·59	1·64	0·98	0·49	0·30	0·24	0·20	0·19	0·18	0·17
	1·74	III	83	10·35	6·34	4·55	2·92	1·71	1·18	0·69	0·46	0·39	0·33	0·31	0·30	0·29
	1·18	III	83	0·86	0·81	0·96	1·30	1·69	1·76	1·57	1·32	1·20	1·10	1·05	1·03	1·01
exponential	1·09	III	83	16·98	10·93	7·57	4·22	1·92	1·13	0·54	0·30	0·23	0·20	0·18	0·17	0·16
	0·87	III	83	6·79	5·12	4·09	2·92	1·94	1·48	0·97	0·68	0·57	0·49	0·45	0·43	0·42
Gaussian	2·18	III	83	9·42	7·47	6·00	4·06	2·28	1·59	0·98	0·62	0·49	0·38	0·34	0·32	0·31
pherical well	3·49	I	150	3·41	—	1·35	0·50	0·11	0·08	0·09	0·31	0·37	1·94	5·34	—	13·63
	3·05	I	150	2·24	—	—	0·54	0·15	0·11	0·25	0·44	0·68	2·16	—	—	8·80
	2·80	I	150	1·66	1·26	0·95	0·56	0·26	0·21	0·37	0·62	0·92	2·15	3·71	4·92	6·55
	2·62	I	150	1·36	1·09	0·88	0·61	0·38	0·36	0·54	0·79	1·06	2·15	3·07	3·89	4·97
	2·18	I	150	1·84	1·78	1·65	1·50	1·43	1·51	1·72	1·67	1·69	2·12	2·73	3·18	3·76
ukawa	2·18	I	150	4·97	1·49	0·74	0·32	0·13	—	0·11	—	0·46	1·25	2·96	6·02	19·84
	1·74	I	150	2·64	1·11	0·64	0·32	0·16	0·12	0·16	0·33	0·57	1·33	2·68	4·60	10·75
	1·18	I	150	0·82	0·51	0·39	0·30	0·26	0·27	0·38	0·61	0·83	1·19	1·60	2·08	3·21
xponential	1·09	I	150	3·66	1·69	0·91	0·36	0·11	—	0·08	—	0·47	1·50	3·70	6·77	14·65
	0·87	I	150	1·92	1·14	0·75	0·39	0·19	0·14	0·18	0·40	0·71	1·61	3·02	4·56	7·61
aussian	2·18	I	150	2·30	1·50	0·99	0·47	0·16	0·10	0·13	0·36	0·69	1·92	3·95	5·93	9·09

herical well	3·49	III	150	17·63	—	6·99	2·72	0·80	0·64	0·15	0·002	0·015	0·025	0·026	—	0·024
	3·05	III	150	9·42	—	—	2·63	1·34	1·10	0·55	0·101	0·021	0·000	—	—	0·009
	2·80	III	150	5·71	4·32	3·35	2·26	1·67	1·77	1·03	0·37	0·16	0·05	0·03	0·02	0·01
	2·62	III	150	3·11	2·47	2·07	1·73	1·84	2·03	1·61	0·77	0·44	0·03	0·02	0·02	0·01
	2·18	III	150	2·02	1·77	1·71	1·94	3·11	4·28	5·14	4·23	3·48	2·71	2·34	2·17	1·77
ukawa	2·18	III	150	23·67	7·62	3·94	1·83	0·76	—	0·20	—	0·09	0·08	0·07	0·07	0·06
	1·74	III	150	10·35	4·78	3·12	1·85	0·97	0·62	0·33	0·21	0·17	0·14	0·13	0·13	0·12
	1·18	III	150	0·86	0·93	1·25	1·65	1·72	1·51	1·09	0·80	0·70	0·62	0·58	0·56	0·55
xponential	1·09	III	150	16·98	8·04	4·62	2·15	0·86	—	0·20	—	0·07	0·06	0·05	0·05	0·05
	0·87	III	150	6·79	4·24	3·07	2·05	1·28	0·89	0·48	0·29	0·23	0·18	0·17	0·16	0·15
aussian	2·18	III	150	9·42	6·23	4·32	2·47	1·34	0·88	0·38	0·17	0·12	0·08	0·07	0·07	0·06
herical well	3·49	I	250	3·43	1·58	0·68	0·12	0·06	0·02	0·01	0·08	0·28	0·52	2·71	6·30	13·66
	3·05	I	250	2·18	—	0·67	0·23	0·14	0·10	0·004	0·29	0·48	0·94	2·71	—	8·75
	2·80	I	250	1·57	0·97	0·61	0·30	0·21	0·13	0·05	0·50	0·71	1·18	2·51	4·00	6·46
	2·62	I	250	1·12	0·74	0·52	0·32	0·25	0·16	0·13	0·70	0·89	1·27	2·18	3·15	4·75
	2·18	I	250	0·87	0·67	0·55	0·45	0·46	0·55	1·08	1·91	1·93	1·81	2·15	2·61	3·39
ukawa	2·18	I	250	4·96	0·89	0·39	0·15	0·06	0·04	0·05	0·11	0·21	0·61	1·59	3·63	19·88
	1·74	I	250	2·66	0·76	0·38	0·17	0·08	0·06	0·07	0·16	0·29	0·75	1·66	3·18	10·79
	1·18	I	250	0·81	0·41	0·30	0·22	0·17	0·17	0·23	0·42	0·63	1·00	1·38	1·82	3·30
xponential	1·09	I	250	3·65	1·06	0·48	0·15	0·04	0·02	0·03	0·08	0·19	0·67	1·98	4·29	14·61
	0·87	I	250	1·91	0·85	0·47	0·21	0·09	0·06	0·08	0·19	0·38	0·94	1·98	3·47	7·64
aussian	2·18	I	250	2·28	1·13	0·58	0·22	0·07	0·03	0·03	0·15	0·33	0·94	2·35	4·52	9·09
herical well	3·49	III	250	17·63	8·21	3·70	1·01	0·57	0·15	0·002	0·024	0·003	0·011	0·015	0·016	0·015
	3·05	III	250	9·42	—	3·18	1·54	1·03	0·55	0·01	0·02	0·01	0·004	0·002	—	0·004
	2·80	III	250	5·71	3·64	2·53	1·75	1·54	1·03	0·12	0·003	0·014	0·010	0·01	0·01	0·007
	2·62	III	250	3·11	2·18	1·81	1·76	2·03	1·61	0·36	0·01	0·004	0·01	0·01	0·01	0·01
	2·18	III	250	2·02	1·71	1·82	2·68	4·58	5·14	3·23	1·05	0·45	0·15	0·07	0·04	0·02
ukawa	2·18	III	250	23·67	4·72	2·26	0·96	0·37	0·20	0·09	0·05	0·04	0·03	0·03	0·03	0·03
	1·74	III	250	10·35	3·52	2·14	1·17	0·55	0·33	0·16	0·10	0·08	0·07	0·06	0·06	0·06
	1·18	III	250	0·86	1·14	1·56	1·76	1·43	1·09	0·67	0·46	0·39	0·33	0·31	0·30	0·29
xponential	1·09	III	250	16·98	5·26	2·68	1·11	0·40	0·20	0·07	0·03	0·02	0·02	0·02	0·02	0·01
	0·87	III	250	6·79	3·37	2·29	1·47	0·79	0·48	0·21	0·11	0·09	0·07	0·06	0·06	0·05
aussian	2·18	III	250	9·42	4·86	2·83	1·58	0·77	0·38	0·10	0·04	0·03	0·02	0·02	0·01	0·01

In conclusion, then, the angular distribution enables considerable light to be thrown on the exchange properties of the nucleonic interaction, but useful information about the state of the potential can only be obtained from quite accurate measurements of the distribution, particularly near scattering angles of 0 and 180°.

(c) *The total collision cross-section*

Table 6 shows the results of the calculation of the variation of total cross-section for *n-p* scattering with incident neutron energy for interactions of types I and II and the different assumptions about the potential shapes. The wide range of variation of the calculated total cross-section and the differences in the form of its variation with incident neutron energy makes it evident that the total cross-section provides a better means of distinguishing between different forms of the potential than does the angular distribution.

TABLE 6

shape of well	range of interaction ($\times 10^{-13}$ cm.)	type of interaction	total collision cross-section ($\times 10^{-26}$ cm.^{2}) for neutron energies in laboratory system (MeV)				
			83	100	150	200	250
spherical well	3·49	I	21·42	—	12·35	—	7·44
	3·05	I	19·80	—	11·56	—	7·43
	2·80	I	19·05	17·04	11·53	9·095	7·71
	2·62	I	18·66	16·82	12·23	9·33	7·97
	2·18	I	25·46	25·51	22·06	17·83	15·32
Yukawa	2·18	I	16·37	—	10·16	—	5·96
	1·74	I	13·97	11·90	8·51	6·65	5·49
	1·18	I	9·84	9·10	7·65	6·62	5·86
exponential	1·09	I	17·94	—	10·29	—	6·24
	0·87	I	14·71	12·62	9·10	7·15	5·90
Gaussian	2·18	I	17·65	14·87	10·35	7·98	6·49
spherical well	3·49	II and III	25·25	—	14·05	—	8·49
	3·05	II and III	23·18	—	13·38	—	8·10
	2·80	II and III	23·32	20·73	14·53	10·82	8·74
	2·62	II and III	23·86	22·16	16·16	12·50	10·02
	2·18	II and III	45·38	48·86	46·02	38·41	30·79
Yukawa	2·18	II and III	19·15	—	11·16	—	6·60
	1·74	II and III	16·12	13·79	9·89	7·73	6·37
	1·18	II and III	17·55	16·49	13·92	11·86	10·35
exponential	1·09	II and III	19·71	—	11·41	—	7·03
	0·87	II and III	16·77	15·27	11·09	8·68	7·13
Gaussian	2·18	II and III	20·81	17·66	12·23	9·28	7·48

Of the four shapes of potential considered in these calculations the spherical well gives the largest values of the total cross-section, the Yukawa potential the smallest. This may be due to some extent to the use of Born's approximation which, as is shown by table 3, will tend to give too large a value of the total cross-section in the spherical well case at incident neutron energies in the neighbourhood of 100 MeV. However, it is evident from the curves that this same behaviour of the total cross-section persists even for neutron energies of 250 MeV where Born's approximation would be expected to agree fairly closely with the accurate calculation, even for the spherical well.

The curves of figure 6 show that the range of the interaction determines to a considerable extent the form of the variation of the total cross-section with incident neutron energy. The decrease of total cross-section with increase of neutron energy

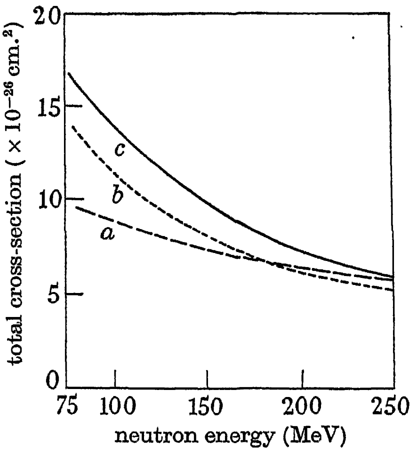


FIGURE 6. Variation of total cross-section with neutron energy for Yukawa potential of three different ranges. Type I interaction. Ranges: *a*, $1{\cdot}18 \times 10^{-13}$ cm.; *b*, $1{\cdot}74 \times 10^{-13}$ cm.; *c*, $2{\cdot}18 \times 10^{-13}$ cm.

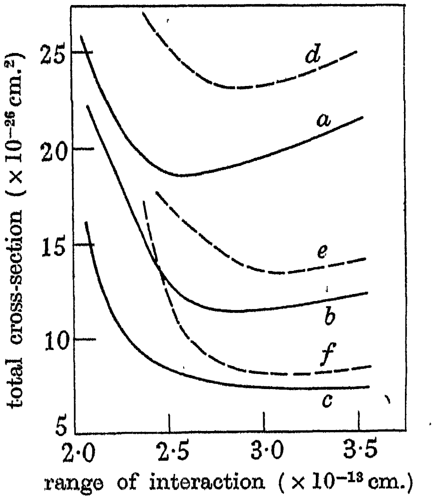


FIGURE 7. Variation of total cross-section with range of interaction for spherical well shape of interaction.

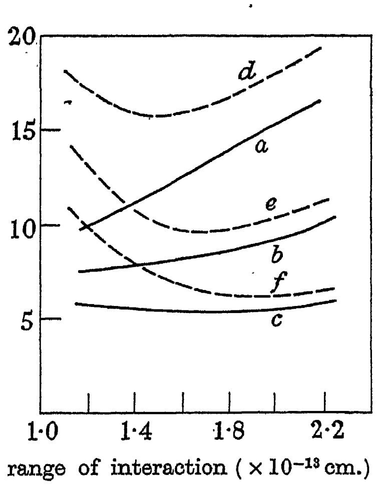


FIGURE 8. Variation of total cross-section with range of interaction for Yukawa well shape of interaction.

Curves: *a*, type I interaction, 83 MeV neutron energy.
b, type I interaction, 150 MeV neutron energy.
c, type I interaction, 250 MeV neutron energy.
d, type II or III interaction, 83 MeV neutron energy.
e, type II or III interaction, 150 MeV neutron energy.
f, type II or III interaction, 250 MeV neutron energy.

is less rapid with the shorter range interactions where the non-central part of the interaction increases in importance.

For a given interaction, a given shape of potential, and a given incident neutron energy, figures 7 and 8 show the variation of total cross-section with the range of the interaction. As the range of interaction decreases the total cross-section passes through a minimum and then increases. This increase at very short ranges is associated with the increasing importance of the non-central part of the interaction. The curves of figures 7 and 8 show that for both the spherical well and the Yukawa potentials, the minimum in the total cross-section-range curves occurs for larger ranges at higher incident neutron energies.

Comparison with experiment

Cook, McMillan, Peterson & Sewell (1947) have reported results of the measurement of the total cross-section for *n-p* scattering using neutrons of energy 90 ± 13 MeV. They obtained the value $(8\cdot3 \pm 0\cdot4) \times 10^{-26}$ cm.^{2}. This is smaller than the value given for this energy for any of the potential shapes investigated in these calculations. Of all the potentials investigated the Yukawa potential of range $1\cdot18 \times 10^{-13}$ cm. and type I interaction gives a total cross-section closest to that observed experimentally. The value given by it is $9\cdot5 \times 10^{-26}$ cm.^{2} for neutrons of energy 90 MeV.* If, as in the case of the spherical well potential, Born's approximation gives too high a value also for the Yukawa potential, the accurate cross-section for this potential would probably lie still closer to the observed value.

For the angular distribution of neutrons of energy about 90 MeV scattered by protons, Hadley, Leith, York, Kelly & Wiegand (1948) have given some preliminary results. It has been shown by Massey *et al.* (1948) that these results could probably be explained on the assumption of a spherical well potential of range $2\cdot80 \times 10^{-13}$ cm. and an interaction of type II. Such an interaction, however, gives much too high a value of the total collision cross-section ($15\cdot8 \times 10^{-26}$ cm.^{2}). The Berkeley experiments gave for the ratio, of the differential cross-section at 180° to that at 90°, a value about 3. The value obtained for this ratio for a type I Yukawa interaction of range $1\cdot18 \times 10^{-13}$ cm. is about 6 (see figure 3). However, owing to the sharp increase in the differential cross-section at scattering angles near 180° for potentials of this type, it seems likely that, unless the measurements near a scattering angle of 180° were carried out with considerable accuracy, a value of the ratio $\sigma_{180^\circ}/\sigma_{90^\circ}$ less than 6 might be obtained.

We are very much indebted to Professor H. S. W. Massey and Dr Tsi-Ming Hu for allowing us to use, prior to publication, the interaction constants calculated by them. We are also indebted to them for many valuable discussions during the course of the work. One of us (H.N.Y.) wishes to express his indebtedness to the Government of Bihar (India) for the grant of a scholarship enabling him to carry out this work.

---

* It should be borne in mind, however, that the interaction constants for this range of interaction and the Yukawa potential were probably calculated rather differently from the constants for other interaction types. The interaction constants for the other cases are those of Hu & Massey (1949). Those for the Yukawa potential of range $1\cdot18 \times 10^{-13}$ cm. are derived from constants given by Chew & Goldberger (1948), and great accuracy was not claimed for them.

REFERENCES

Ashkin, J. & Wu, T. Y. 1948 *Phys. Rev.* **73**, 973.
Barker, F. C. 1948 *Nature*, **161**, 726.
Camac, M. & Bethe, H. A. 1948 *Phys. Rev.* **73**, 191.
Chew, G. F. & Goldberger, M. L. 1948 *Phys. Rev.* **73**, 1408.
Cook, L. J., McMillan, E. M., Peterson, J. M. & Sewell, D. C. 1947 *Phys. Rev.* **72**, 1264.
Eisenstein, J. & Rohrlich, F. 1948 *Phys. Rev.* **73**, 641, 1411.
Hadley, J., Leith, C., York, H., Kelly, E. L. & Wiegand, C. 1948 *Phys. Rev.* **73**, 541.
Hu, T. M. & Massey, H. S. W. 1949 *Proc. Roy. Soc.* A, **196**, 135.
Massey, H. S. W., Burhop, E. H. S. & Hu, T. M. 1948 *Phys. Rev.* **73**, 1402.
Rarita, W. & Schwinger, J. 1941 *Phys. Rev.* **59**, 436, 556.

Recent advances in the technique of submarine gravity surveying

BY R. I. B. COOPER

Department of Geodesy and Geophysics, University of Cambridge

(*Communicated by Sir Gerald Lenox-Conyngham, F.R.S.— Received* 11 *December* 1948)

The errors in gravity surveying at sea which are the most difficult to estimate are involved in position-finding, measurement of speed over the ground, the variations of depth-keeping and in the second-order correction. Submarine gravity surveying is also a slow process on account of the time required for an observation with the standard Vening Meinesz 3-pendulum apparatus and the complications of computing the result. During the recent cruise in the English Channel in May 1948 of H.M. Submarine *Talent* most of these difficulties were overcome by the provision of special instruments which are described here. The most important of these was a stabilizer which, by automatically compensating the effect of sideways accelerations on the pendulums, removed the necessity for certain tedious corrections and enabled the results to be rapidly computed.

1. INTRODUCTION

The technique of making pendulum observations at sea was introduced first by Meinesz in 1923 (Meinesz 1929). Since then many expeditions have been made by various nations employing the apparatus in the form developed by him. Ewing made important advances (Ewing 1937) by employing the quartz crystal as a standard of frequency for timing the pendulums and by eliminating the variable lag in the electromagnetic shutter which, by eclipsing the light source, provided timing marks on the record. His technique was employed almost unchanged by Browne (1937) and by Browne & Cooper (1949) in later British surveys.

Browne (1937) made an important theoretical contribution to the problem by pointing out that the accelerations of the submarine, though compensated to the first order by the 3-pendulum technique, would give rise to second-order effects. No practical confirmation of this theory has yet been obtained, though attempts to do so have been made by Meinesz (1937, 1938) and by Browne & Cooper (1949).

The latter obtained no results confirming the theory but believed that other sources of error, e.g. vibration of the pendulum support and uncertainties in ground speed, may have masked the effect for which they were seeking.

One of the objectives of the expedition in H.M. Submarine *Talent* in May 1948 in the English Channel was the investigation of these sources of error. By the use of a 'Decca' navigation system which provides excellent cover in the Channel, and of taut-wire gear for ground-speed measurement, two of the main uncertainties were eliminated. The estimated accuracy of position-finding was about 100 yd., equivalent to 0·05 mgal. in a north to south direction, and that of ground speed was about 0·05 knot, equivalent to 0·2 mgal. in an east to west direction.

Variations in depth-keeping have the effect of bringing about a difference of vertical velocity between beginning and end of an observation and hence of causing a change in the mean value of apparent gravity taken over the whole time. If a limit can be safely assigned to the maximum differences in vertical velocity, then by observing for a sufficiently long time the average vertical acceleration can be reduced to any given amount. In practice, observations of at least half an hour are essential at present. Part of the velocity variations are, of course, due to wave motion, but the nature of the Meinesz record enables these short-term fluctuations to be averaged out. It is the long-term variations which have to be taken into account.

In *Talent* an automatic depth recorder was employed, producing a record of depth variations accurate to about 1 ft. with respect to any previously arranged zero. The record was obtained in permanent form from an Esterline-Angus recording milliammeter, and changes in vertical velocity of about 0·5 ft./min. could be measured from it. This enabled the error from this cause in observations of 20 min. duration to be reduced if necessary to 0·04 mgal. Corrections of as much as 2·5 mgal. have sometimes been derived from the records. Shorter observations could be taken but for the limitations of the accuracy of recording the pendulum swings.

The acceleration recorder described by Browne & Cooper was carried in almost unchanged form. It was expected that, other errors having been substantially reduced by the above refinements, any remaining discrepancies between two or more observations at different depths at the same station should be capable of close correlation with the measured accelerations.

A serious disadvantage of the pendulum technique is its slowness. This is due, first, to the necessity for developing a photographic record, which cannot be undertaken at sea in a submarine and so must be postponed, and secondly, to the length of time required to compute the corrections which have to be applied before the value of gravity becomes known. It is clear that an automatic method of recording the motion of the fictitious pendulum would be desirable, and although any real system would have to take account of the fluctuations in period due to the vertical accelerations of the wave movements, it is, nevertheless, possible to conceive one without great difficulty. The advantage would be lost, however, if some of the corrections could not be eliminated. The most tedious of these are those for arc and for deviation from isochronism, depending as they do on a record of the amplitude and phase of the middle pendulum with respect to the two fictitious ones for the whole observing time.

In *Talent*, therefore, an instrument was carried which stabilized the pendulum apparatus against horizontal accelerations in the swinging plane, and kept the centre pendulum at rest throughout. This was done by arranging that the displacement of the centre pendulum from the apparent instantaneous vertical controlled, by means of a photocell and electrical amplifier, the current in an electromagnet rigidly attached to the submarine and attracting an armature secured to the frame of the apparatus. Since all three pendulums are mounted on the same support inside the apparatus, if one of them remains undisturbed and at rest, whatever the motion of the support may happen to be, then it can be shown that, subject to certain conditions, the other two will swing with the same period as they would possess on a rigid support.

The deviation from isochronism correction is thus eliminated, and the arc correction is also simplified.

This 'stabilizer' has been shown theoretically and practically to be successful and has enabled the results of the *Talent* survey to be computed very quickly. These results and their geological significance will be discussed in a later paper.

During the cruise, such fine weather was experienced that the second-order corrections even on shallow dives were smaller than the probable error of uncorrected gravity. Therefore no information has been obtained at sea either confirming or conflicting with Browne's theory. After return from the cruise an attempt was accordingly made to apply artificial accelerations to the entire apparatus in the laboratory. Though these experiments were incomplete because the pendulum apparatus was needed elsewhere, they confirmed the theory for accelerations of 5 sec. period, but anomalous results were obtained when the period was reduced to 3 sec.

The report describes the theory and practice of the stabilizer, the construction of the automatic depth recorder, and the use of the navigational aids. Results are introduced only so far as they are relevant to this discussion, and the full table of gravity values will be given in the later paper.

THE STABILIZER

2. THEORY OF THE ACTION OF THE STABILIZER

If a single pendulum is hung on an immovable support it is a relatively easy matter to deduce the changes in gravity from the changes in its period when it is set up in different places. Most supports are elastic, however, and a correction has to be applied to allow for the periodic yielding to the pull of the swinging pendulum. Alternatively this 'sway' correction can be eliminated by swinging two isochronous pendulums with equal amplitude in antiphase on the same support. If the support is now given a random motion it is clear that the swinging of the pendulums is disturbed, and that if the motion is not completely specified no result can be deduced for the value of gravity. Meinesz (1929) has shown, however, that the difference between the deflexions of two isochronous pendulums swinging on the same support is unaffected to the first order, provided the motion of the support contains only periods which are long compared with that of the pendulums. This differential oscillation he terms a fictitious pendulum. If the pendulums are not exactly isochronous this result still holds if the deviation from isochronism is small, provided that a correction is applied for it. It is clear that, even were the pendulums isochronous at zero amplitude, since their individual amplitudes vary with time depending on the motion applied to the support, they will not remain so throughout the observations. Thus there is also an arc correction of complicated form to be applied. The corrections for arc and deviation from isochronism (abbr. DI) appear in the following form:

$$\text{arc} \quad \delta T_1 = +\tfrac{1}{16}T_1\{a^2+a_2^2+3aa_2\cos(\phi_2-\phi)+2a_2^2\cos^2(\phi_2-\phi)\}, \tag{2·1}$$

deviation from isochronism

$$\delta T_2 = (T_2-T_1)\,a_2\cos(\phi_2-\phi)/a, \tag{2·2}$$

where T_1, T_2 = periods of pendulums 1, 2; a, a_2 = amplitudes of fictitious pendulum and of pendulum 2; and ϕ, ϕ_2 = phases of the same.

The term 'fictitious pendulum' is used to describe a movement representing the difference between pendulums 1 and 2.

It will be seen that, since a_2 and $\phi_2-\phi$ are not constant, their values will have to be read from the record and averaged over the whole observation. This process absorbs most of the time of computing. If a_2 is small there is some advantage, since the accuracy with which the other quantities must be determined is thereby reduced. On the other hand, if it is arranged that initially one pendulum is at rest while the other swings there will be a sway correction; accordingly three pendulums are used, of which the outer two are set swinging in antiphase whilst the middle one is initially at rest. Records are made of the two fictitious pendulums formed by the difference of the motions of the middle and each outer pendulum. Although this secures absence of sway at first, during the course of an observation the system may change from what we may call the $(1, 0, -1)$ state to a $(2, 1, 0)$ state, thereby bringing about a large sway correction. If the motion of the support is fairly small and random, however, the sway will average itself out over the time of observation. Furthermore, the sway correction will be of opposite sign for the two fictitious pendulums, since these continue to swing in antiphase. Thus to the first order their mean period, which is always taken for deducing gravity, is unaffected. This effect is, however, shown by variations in the differences of their periods, which, at sea, may amount to as much as 50×10^{-7} sec.

Nevertheless, it is clear that considerable advantages of convenience would result if a_2 were always zero. It is easily seen from (2·1) that the arc correction then reduces to the usual expression except that the amplitude of the fictitious pendulum appears; Meinesz has shown that this also is unaffected by the motion of the support, and so a single observation of this amplitude at the beginning and another at the end would suffice for correction purposes, as is done for land observations. The correction for deviation from isochronism vanishes.

There are several complications which arise when an auxiliary electromechanical device (or 'servo-stabilizer') is considered for this purpose. In the first place it is clear that the condition $a_2 = 0$ implies that the support must in fact remain stationary as regards horizontal motion in the swinging plane. Any movement whatever will destroy the condition. This is clearly an absurd restriction on an apparatus mounted in a submarine. The difficulty is, however, only apparent; the condition $a_2 = 0$ really implies that the middle pendulum does not depart from the apparent vertical, not the true vertical. It is clear that if the direction of apparent gravity were varied infinitely slowly the outer pendulums would swing with unchanged period about an inclined zero line, whilst the centre pendulum would aline itself parallel to this zero line; the fictitious motion would be undisturbed. In the apparatus is mounted a damped pendulum of short period which is intended to maintain its alinément along the apparent vertical; it is therefore necessary only to arrange for the middle pendulum to remain parallel to this damped pendulum to keep $a_2 = 0$.

Even if this condition is satisfied exactly—although it evidently cannot be in practice because of the technical imperfections considered later—it must be borne in mind that the pendulums are not exactly isochronous even at zero amplitude. Accordingly it is not evidént without further investigation that the outer pendulums will swing as if undisturbed. Putting the problem another way, we may enquire, given that energy in a band centred on the middle pendulum period is filtered from the support, how much energy will be acquired by the other pendulums at periods slightly above or below this.

Put in the latter form, the problem can be answered by a very simple argument. Regard the system whereby the middle pendulum is reduced to rest as effectively damping the latter artificially and rather heavily, depending on the technical efficiency attained. Then its response curve to forcing will be considerably broadened. If its equation of motion is

$$(p^2 + 2\lambda p + n^2)\, y = F, \tag{2·3}$$

where p stands for d/dt and λ is the decrement, then near the resonant point the amplitude response to an acceleration F of constant amplitude and variable frequency may be represented approximately by the formula

$$a/a_0 = 1 - 2\delta n^2/\lambda^2, \tag{2·4}$$

where $\delta n/2\pi$ is the deviation of the frequency of F from resonance. Suppose now that energy of all possible frequencies is contained in the random sideways accelerations which are reaching the whole apparatus. (It will be permissible to suppose the energy uniformly distributed with respect to frequency in a small band within, say, 1 % of $n/2\pi$.) Suppose also that the centre pendulum is kept at rest by the stabilizing device; then no energy of frequency $n/2\pi$ has been allowed to reach the actual support of the pendulums. If the original energy flux were such that after a given interval of time, unit amplitude would have been acquired in the absence of the stabilizer, we may regard the latter as having subtracted energy equivalent to this unit amplitude. At any other frequency $(n \pm \delta n)/2\pi$ the corresponding amount is equivalent to an amplitude of $1 - 2\delta n^2/\lambda^2$, leaving a residual amplitude of $1 - (1 - 2\delta n^2/\lambda^2)$ or $2\delta n^2/\lambda^2$. For this acquired amplitude not to cause an error in apparent period of more than 1 part in 2 million, or 1 mgal. error in gravity, it must not exceed $\pi/500$ of the amplitude of its initial swinging, if the observation lasts 2000 sec. Thus if conditions are such that a free pendulum left on the support would acquire the same amplitude after 2000 sec. as is used when setting the pendulums swinging for an observation—and this corresponds to the roughest conditions ever experienced in practice—we have

$$2\delta n^2/\lambda^2 \ngtr \pi/500. \tag{2·5}$$

Now it will appear later than λ can usually not be made better than $1/15\,\text{sec.}^{-1}$ so that (2·5) reduces to

$$\delta n \ngtr \{\pi/225000\}^{\frac{1}{2}},$$

or

$$\delta T \ngtr 1 \text{ in } 270 \text{ (approx.)}.$$

This is a very easily satisfied condition.

The assumption has been made in this argument that the resultant motions of the pendulums at any time after starting can be regarded as compounded of the original motion plus the linear addition of random vectors; although it is known that a pendulum with finite amplitude is in fact a non-linear system it does not appear that this will invalidate the above argument, which is intended only to give an order of magnitude.

A more serious first-order source of error will be due to the technical impossibility of maintaining a_2 exactly zero under disturbed conditions. Hence, if its value is neglected in equation (2·1) an error will arise. We may easily deduce the root mean square value of a_2 under given conditions if λ is known. If a unit displacement is given to the middle pendulum its response will be

$$e^{-\lambda t} \sin(2\pi t + \alpha), \tag{2·6}$$

where α is an arbitrary phase-angle.

Hence if a series of displacements $a_1, a_2, \ldots, a_n$ are applied at arbitrary intervals, throughout a time of observation T long compared with $1/\lambda$ the root mean square value of a_2 is

$$(a_1^2 + a_2^2 + \ldots + a_n^2)/T \times \int_0^\infty \{e^{-\lambda t} \sin(2\pi t + \phi)\}^2\, dt$$

$$= (a_1^2 + a_2^2 + \ldots + a_n^2)/T \times 1/4\lambda(1 + \lambda^2/4\pi^2). \tag{2·7}$$

Now for a constant energy flux the resultant amplitude is proportional to the square root of the time; as above, it will be taken as equal to the main amplitude after 2000 sec. Therefore under stabilized conditions, if $\lambda = 1/15\,\text{sec.}^{-1}$,

$$\text{root mean square } a_2 = a \times 15^{\frac{1}{2}}/4^{\frac{1}{2}}\, 2000^{\frac{1}{2}} \simeq a/23. \tag{2·8}$$

The total value of the expression (2·1) usually amounts to the equivalent of 12 mgal. Hence, if a_2 does not exceed 5 % of a, the worst error that can occur in the mean of the two fictitious pendulums is 10 % of 12 mgal. or slightly more than a milligal. If the pendulums do not differ in frequency by more than 10 parts per million the error through neglect of DI is also less than a milligal.

Owing to the action of the stabilizer, there will be terms in the motion of the support of period equal to the pendulum period since it cannot be perfectly stabilized. Their amplitude, as has been shown above, will nevertheless be small; in practice the amplitude a of (2·8) is about 0·01 radian, and so about 0·0005 radian is the expected value for the 'imperfection' disturbance term. The general solution to the equation representing the effect of a disturbance of arbitrary period and amplitude on the pendulums has not been obtained. It has been solved in the special case of periods long compared with the pendulum period and for the case of two pendulums swinging in antiphase on the same support by Browne (1937). The equation is

$$l\ddot{\theta}+g\sin\theta = F\cos\theta. \tag{2·9}$$

The general case need not be studied in the present instance, however, since the amplitudes are restricted and the motion of one of the pendulums is constrained in a special manner. Making use of these limitations we proceed as follows: write down the equation for each pendulum in the form (2·9)

$$\begin{aligned} l_1\ddot{\theta}_1+g\sin\theta_1 &= F\cos\theta_1, \\ l_2\ddot{\theta}_2+g\sin\theta_2 &= F\cos\theta_2. \end{aligned} \tag{2·10}$$

F is to be regarded as arbitrary for the present. Eliminating it we obtain

$$(\ddot{\theta}_1-\ddot{\theta}_2)+g\sec\theta_1\sin(\theta_1-\theta_2)/l_2 = (1-l_1\cos\theta_2/l_2\cos\theta_1)\,\ddot{\theta}_1. \tag{2·11}$$

Let θ_1 refer to the stabilized pendulum, then making use of the above limitations we may assume $\theta_1 = \theta+\phi$, where $\tan\phi = F/g$ and θ is small, of order 0·0005, and period 1 sec. Then

$$\begin{aligned} &(\ddot{\theta}_1-\ddot{\theta}_2)+g\sec\phi\sin(\theta_1-\theta_2)/l_2 \\ &\qquad = (1-l_1\cos\theta_2/l_2\cos\theta_1)\,\ddot{\theta}_1+g\sin(\theta_1-\theta_2)(\sec\phi-\sec\theta_1)/l_2. \end{aligned} \tag{2·12}$$

If $\theta = 0$, $l_1 = l_2$ and $\theta_1 = -\theta_2$ we obtain zero on the right-hand side of (2·12) which then reduces to Browne's case. If F is a slow motion then the $\sec\phi$ on the left-hand side may be averaged and we obtain his second-order correction. If $\theta \neq 0$ the effect of the right-hand side regarded as a disturbance term must be assessed. The departure from isochronism is unlikely to affect the case, so we suppose $l_1 = l_2 = l$, and since all the angles are small, we expand the circular functions. Then the right-hand side of (2·12) becomes

$$(\tfrac{1}{2}\theta_2^2-\tfrac{1}{2}\phi^2-\phi\theta)(\ddot{\theta}+\ddot{\phi})+g(\phi-\theta_2)\,\theta\phi/l. \tag{2·13}$$

If we take θ_2 and ϕ to be about 0·01 and the period of ϕ to be about 6 sec. then it is clear that the greatest of these terms is $g(\phi-\theta_2)\,\theta\phi/l$, which may reach an amplitude of 4×10^{-6}. Hence we may rewrite equation (2·11) omitting the right-hand side in the form

$$(\ddot{\theta}_1-\ddot{\theta}_2)+g\sec\phi\sin(\theta_1-\theta_2)(1+\phi\theta)/l = 0. \tag{2·14}$$

By a theorem of Sturm & Liouville, the period of this motion must lie between $(1\pm|\phi\theta|)^{\frac{1}{2}}$ times the period obtaining when $\theta = 0$. Thus if $\phi\theta = 5\times10^{-6}$ it is possible that errors of 5 mgal. might occur. However, it must be appreciated that these conditions represent extremes and smaller errors in fact arise when averaged over a long observation if F is random. Moreover $\phi = 0·01$ would represent seriously disturbed conditions. In the survey in *Talent* described below, ϕ never exceeded an average of 3×10^{-3}, or θ an average of 3×10^{-4}, so that $|\phi\theta|$ was less than 10^{-6}.

3. THEORETICAL DESIGN OF THE STABILIZER

In figure 1 is shown a schematic diagram of the elements of the servo-mechanism. It is assumed that a voltage can be obtained linearly proportional to the difference of inclination $(y-z)$ between the middle main pendulum and the short-period pendulum; if the latter follows the apparent vertical exactly then this represents the quantity α_2 of the previous section. There are two complications: first any real pendulum must have a finite period and so cannot follow the apparent vertical precisely; secondly the short-period pendulum is mounted inside the pendulum case for optical convenience and so is affected by the force applied by the electromechanical transducer. This means that the natural period and damping of the short-period pendulum must enter as parameters in the equations for the system, as must also the natural period and damping of the apparatus itself on its supporting stand. It may be pointed out here that the case is mounted in gimbals to eliminate the effects of the rolling and pitching of the submarine; this will not affect the operation appreciably, since the result is simply that the force of the magnet is not applied quite perpendicular to the direction of apparent gravity.

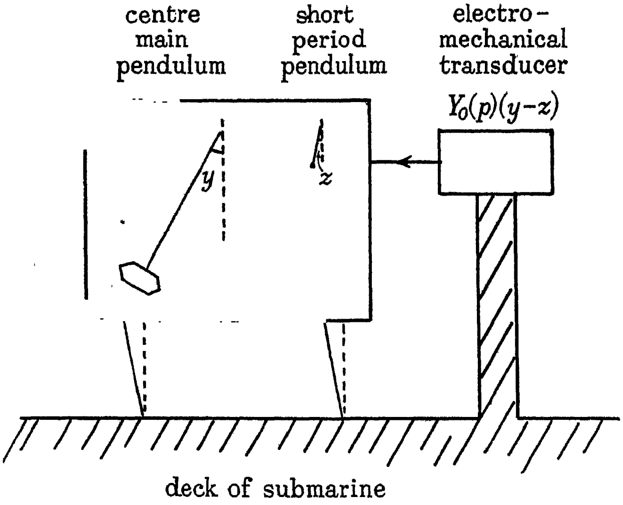


FIGURE 1. Elements of the stabilizer. The transducer applies a force $Y_0(p)$ for a unit input of frequency $p/2\pi$.

The study of such a system, involving the interaction of three separate periodic motions, is a laborious problem, and although it has been carried through and the conditions for stability arrived at, it is clear that the results could only be regarded as pointers and need not be adhered to strictly in the construction. Indeed it will appear that the practical difficulties of modifying the Meinesz pendulum apparatus, which is the only type available, to satisfy the optimum conditions would be prohibitive. Nevertheless a solution will be given for its interest:

$$\left.\begin{gathered} n_1 = (2+\surd 3)^2 n_2, \quad n_3 = (2+\surd 3)\, n_2, \\ \lambda_1 = 0, \quad \lambda_3 = 3n_3, \\ Y_0(p) = 32g/\{(2\surd 3 - 3)\, p/n_3 - 3(7-4\surd 3)\}, \end{gathered}\right\} \tag{3·1}$$

where $n_2 = 2\pi \times$ frequency of main pendulum $= 2\pi\,\text{sec.}^{-1}$,

$n_3 = 2\pi \times$ frequency of short-period pendulum,

$n_1 = 2\pi \times$ frequency of apparatus in mounting,

$\lambda_3 =$ decrement of short-period pendulum,

$\lambda_1 =$ decrement of apparatus in mounting,

$Y_0(p) =$ response of amplifier in Heaviside's notation.

It will be seen that this gives a period for the short-period pendulum of about 1/4 sec. and a heavy damping, and a period for the apparatus of about 1/14 sec., i.e. it should be on a rather rigid mounting, and not damped. The air damping of the main pendulum is, of course, negligible.

When the behaviour of the system under various accelerations of the submarine was studied theoretically, however, it was found that what tended to occur was that forcing of short period caused the short-period pendulum to aline itself along the direction of the middle main pendulum and not *vice versa*. This forcing scarcely disturbed the response of the main pendulum to S, the submarine's acceleration, and so no stabilization took place. This behaviour would probably affect the apparent period of the other main pendulums.

It has already been pointed out that if the short-period pendulum were suspended from a support rigidly mounted in the submarine the optical system for producing a movement equal to its deviation from the middle pendulum mounted in the case in gimbals would be excessively complicated. An amplifier cutting off sharply below 1 cycle would also be very difficult to make. A solution alternative to (3·1) would be $n_1 = (2-\surd 3)\,n_2$ but then the natural period of the apparatus in its mounting (now 4 sec.) would be close to wave-period and dangerous resonances might occur. This difficulty was not overcome, but, as will be seen below, for practical reasons many other criteria had to be abandoned as well and the ultimate test, that the system was made to work and gave results in agreement with unstabilized observations, remains the principal cause for satisfaction.

4. PRACTICAL DESIGN OF THE STABILIZER

It was impracticable to carry out exactly the conditions arrived at theoretically in the last section. The main difficulty was that the existing short-period pendulum in the apparatus had to be used; this was not nearly of short enough period (3/4 sec.) and was insufficiently damped (after considerable patience with adjusting the quantity of oil in the damping compartment the pendulum decayed to e^{-1} of its amplitude in about 12 to 14 swings). This could not well be altered; the pendulum has to carry two reflecting surfaces of approximately 3×1 cm. in area, which in the actual apparatus are heavy prisms, and has to be composite with internal damping (Meinesz 1929, p. 51) to avoid influence by the rolling of the apparatus in its gimbals.

Since three of the most important criteria were then necessarily going to be modified, it was hardly worth endeavouring to design the amplifier according to theory, and the practical procedure adopted was to estimate the force required when manually reducing the pendulum to rest from a certain amplitude, and to build an amplifier with sufficient gain to do the same. A certain amount was kept in hand and a control provided, and the system was adjusted to best stability, as judged by the rate at which the middle pendulum came to rest after an impulse, by experimenting with different amounts of phase-shift.

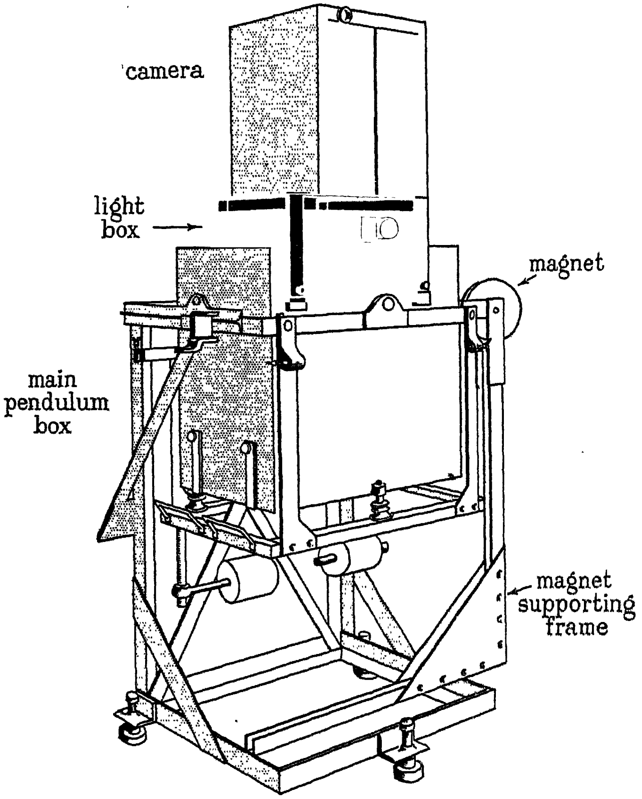


FIGURE 2. Sketch of Meinesz pendulum apparatus with stabilizing attachments.

The rigidity of the apparatus in its mountings also enters into the design; the standard frame provided for a Meinesz apparatus is distinctly 'wobbly' even when clamped to the deck; in *Tudor*[*] it had been considered desirable to brace the upper members to the hull with auxiliary struts to avoid sway. The natural period seems to be about 3 to 4 cycles and this was sufficiently 'high' in comparison with the main pendulum period and that of the 'short-period' pendulum to need no artificial modification.

A crude electromagnet was used as electromechanical transducer. This was mounted rigidly on an L-shaped girder bolted to the lower girders of the frame of the Meinesz apparatus (see figure 2), and its armature was fixed to one of the upper arms supporting the gimbal cradle. The unit is, of course, distinctly non-linear and not very efficient, since the attractive force varies inversely as the fourth power of the air-gap. If a large air-gap is not left instability may arise and the armature can be violently sucked on to the magnet by the large forces developing in the reduced gap with consequent damage to the pendulums. But if too large a gap is left much flux is lost and insufficient mechanical force is obtained.

---

[*] The previous survey, in 1946, was carried out in H.M. Submarine *Tudor*.

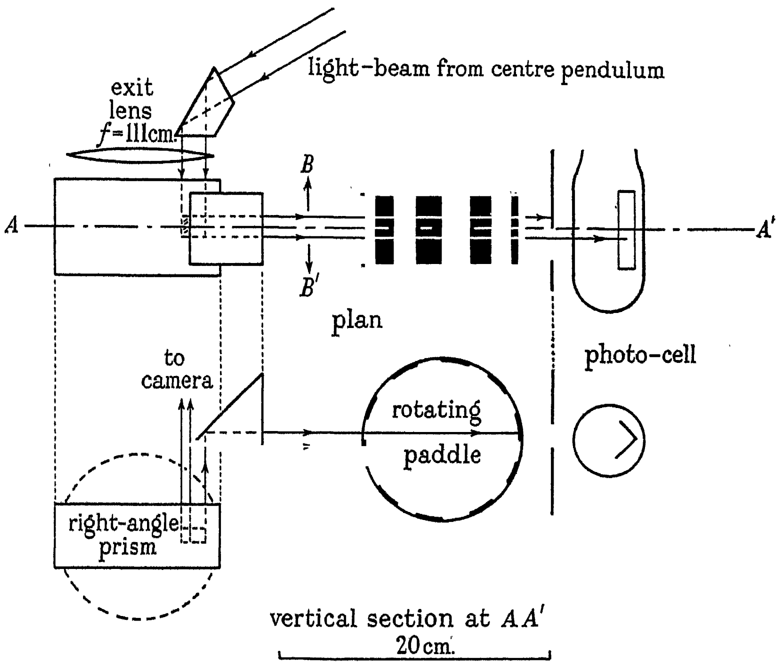


FIGURE 3. Schematic diagram of optical system. BB' = direction of beam movement.

Although it thus appears that the practical realization of the theoretical design left much to be desired, it must be borne in mind that the properly designed theoretical instrument would have been nearly critically damped; all that was necessary for the immediate ends was a device for abstracting energy from the middle pendulum at least as fast as it would be acquired; this meant that it should decay in a few hundred swings, not in one or two. In the final instrument about 20 swings were required before the amplitude became inappreciable, so that there was still an adequate safety margin, even in the relatively crude design necessitated by practical limitations.

The relative motion of short-period pendulum and middle pendulum had to be made to produce a proportionately varying current from a photocell; it was also clearly essential to retain a photographic record of the movement of the beam in order to be able to check afterwards that the stabilizer had worked properly. It was impossible to arrange for two separate light sources and so the light beam was split after emergence from the pendulum apparatus (figure 2, lower box) on its way to the camera (figure 2, upper box) by inserting the edge of a right-angle prism. Figure 3 shows a schematic diagram of the optical system; it was housed in the small black box seen in figure 2 just below the camera. A stop had also to be inserted in the train of prisms inside the pendulum apparatus where the light is parallel to prevent rolling of the apparatus in its gimbals from affecting the amount of light falling on the photocell.

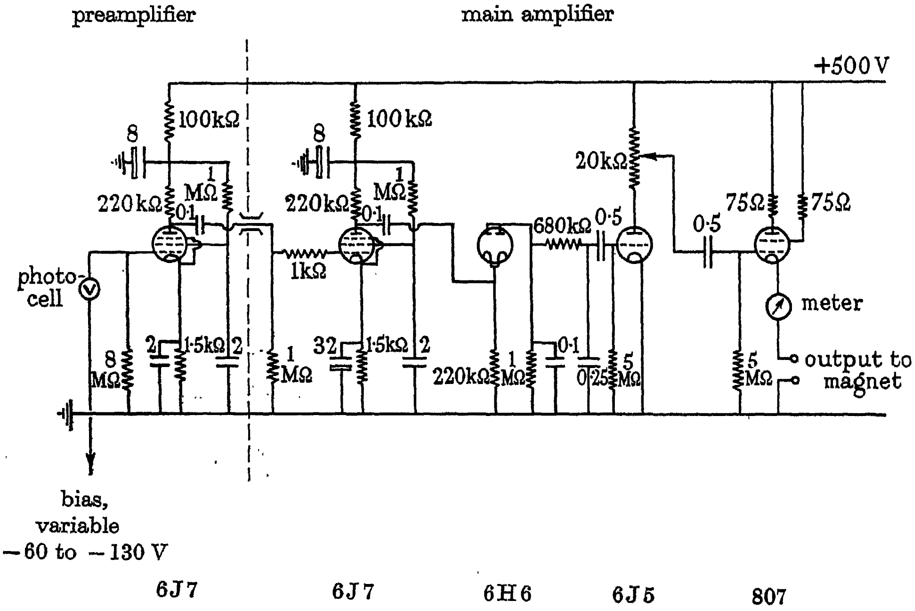


FIGURE 4. Circuit diagram of stabilizer amplifier.

An a.c. amplifier was employed to amplify the current from the latter in order to secure sufficient gain conveniently, and so the light was 'chopped' by a rotating paddle-wheel before it fell on the cell. This can also be seen in figure 3. A d.c. Pointolite was used as light-source to prevent 50 c./sec. hum from causing interference. Considerable care was taken to screen the leads to the motor driving the paddle-wheel. A preamplifier valve was mounted in the box as this reduces noise and interference by raising the signal-level in the lead from box to amplifier; the separation of the two stages of amplification also makes for stability. A circuit diagram of the system is given in figure 4.

In the normal operation of the Meinesz apparatus, an alternative method of recording is adopted during the middle of the observation. The paper is run at one-third the normal speed and the light is eclipsed for half a second every second. This enables the phases of the centre pendulum and fictitious pendulums to be compared during this time. In the present instance it is not possible to do this since the eclipses will affect the light falling on the photocell and disturb the apparatus. However, since the phases are not required when the stabilizer is working, this drawback is not serious.

The following numerical data give an idea of the operation:

Photocell sensitivity, μA cell-current at optimum bias per mm. movement of spot on record	10^{-4}
Grid leak in series with photocell	10 MΩ
Gain per stage at 250 c./sec.	120
Voltage (d.c.) at grid of 6J 5, per mm.	0·6
Magnet current at ½-gain, per mm.	3 mA
Magnet pull, per mA (gap = 2·5 mm.)	300 g.
Overall sensitivity, g./mm. at ½-gain	900

Magnet wound with 15,000 turns no. 27 enamelled copper on 1 in. diam. iron core 2 in. wide, normal standing current 60 mA, pole-face area 10 cm.^{2}.

5. RESULTS OF TESTS ON THE STABILIZER

There was insufficient time to design a proper shaking table before the cruise. All that could be done in the way of tests was to take runs alternately with the stabilizer switched on and off and to apply deliberate disturbances by pushing the frame during the stabilized runs. Even without these the stabilizer can be tested, because in general the two outer pendulums are not exactly balanced and some sway is present which causes the centre pendulum to acquire amplitude unless the stabilizer is on.

When adjusting the system at first to get the centre pendulum to come to rest in the shortest number of swings it was noticed that the amplitude often decayed quite rapidly in a few swings and then built up again only to decay once more, a series of 'beats' being superposed on the general decay. This was attributed to the closeness of the periods of centre pendulum and short-period pendulum and to the poor damping of the latter; probably when the first apparent zero was reached both pendulums were swinging in phase with equal amplitude, i.e. the apparatus had been 'pushed' so as to bring the short-period pendulum into line with the centre pendulum and not *vice versa*. The possibility of this occurring when the short-period pendulum is mounted in the apparatus and so affected by the artificial stabilizing impulses has been mentioned in the theoretical section. Damping the short-period pendulum by external means (tissue paper) improved the performance considerably, but in the case of rolling of the apparatus at sea, this would cause disturbance of the pendulum. It seems certain that this is what limited the best rate of decay to about 15 to 20 swings.

It was found that during a stabilized swing, even on the rigid floor of the laboratory, the centre pendulum retained a small amplitude, less than 0·0005 radian. It was at first thought to be due to some instability in the system, but when the main pendulums were at rest prior to starting the centre pendulum could be brought completely to rest. It therefore appeared that the amplitude represented the response to forcing by the inevitable small lack of symmetry in starting the pendulums. This suggested the following experiment: one pendulum alone was swung with normal amplitude and the initial rate of gain of amplitude of the centre pendulum noted when the stabilizer was switched off. This was found to be 0·01 radian in about 3 min.

The stabilizer was then switched on and the final steady amplitude of the centre pendulum noted; this was about 0·0006 radian. Now the normal maximum rate of gain under submarine conditions is 0·01 radian in 30 min. and since the amplitude gained is proportional to the square root of the time, $\sqrt{10}$ times as great an amplitude would have been gained in this time in our experiment. Therefore we expect under submarine conditions only $1/\sqrt{10}$ of the equivalent steady amplitude, i.e. 0·0002 radian. This is less than the steady amplitude due to unbalance and both are in any case negligible.

TABLE 1. LABORATORY TESTS ON STABILIZER (AT PENDULUM HOUSE)

obs. no.	A (cm.)	B (cm.)	pendulum difference (10^{-7} sec.)	mean period 0·50
S 1	—	—	91,2	3,9073,8
S 2	0·16	4	94,5	3,9073,7
S 3	—	—	92,8	3,9076,7
S 4	0·30	7	83,6	3,9072,0
S 5	—	—	92,3	3,9074,0

A = root mean square amplitude.
B = final (calculated) amplitude if stabilizer had not been in operation.
[Pendulums 2 and 3 in use.]

The results of the artificial disturbance tests are given in table 1. In the case of observations S 2 and 4, a bump was given to the apparatus every minute, and the stabilizer allowed to reduce the amplitude thus acquired as far as it was able before the next impulse. The equivalent root mean square amplitudes were estimated from the record. No correction was applied to allow for the amplitude of the centre pendulum, the arc corrections being simply estimated from a single measurement in the middle of each record of the two fictitious pendulums. The observations S 1, 3 and 5, were undisturbed but were taken with the stabilizer switched on so that effectively the centre pendulum was at rest throughout. It will be seen that the differences are quite permissible in view of other errors which would enter at a sea station.

Tables 2 and 3 show the initial and final base-station swings, each consisting of six observations, three stabilized and three unstabilized. Although the periods changed during the voyage the difference on each occasion between the mean stabilized and unstabilized swings are not significant.

No particular difficulty was experienced in installing and operating the stabilizer in the submarine. Harbour swings were made successfully at Portland from 28 to 30 April and, as will be seen from table 4, both unstabilized and stabilized swings agreed with one another, although on account of the exceptionally calm weather not much amplitude was gained by the centre pendulum even when unstabilized, so that this was not a very severe test.

Once set up, the stabilizer gave no trouble and functioned as intended throughout the voyage. Adjustment of the bias on the photocell to give optimum sensitivity was made by observing the a.c. output from the second amplifier stage on an oscillograph.

[The latter had to be carried in any case for general test purposes.] The bias was not stabilized and required occasional readjustment. This rather inconvenient experimental procedure was simply dictated by lack of time to build the obviously indicated modifications. It did not prove a source of annoyance.

TABLE 2. INITIAL BASE STATION SWINGS WITH AND WITHOUT STABILIZER (AT PENDULUM HOUSE)

obs. no.	stab.	pendulum difference (10^{-7} sec.)	mean period 0·50
FB 2	off	47,6	3,9048,6
3	on	55,6	3,9051,4
4	off	50,3	3,9048,7
5	on	52,3	3,9051,2
6	off	50,5	3,9050,5
7	on	45,9	3,9048,7

Mean pendulum periods, stabilized = 0·503,9050,4.
Mean pendulum periods, unstabilized = 0·503,9049,3.
[Pendulumns 1 and 3 in use.]

TABLE 3. FINAL BASE STATION SWINGS WITH AND WITHOUT STABILIZER (AT PENDULUM HOUSE)

obs. no.	stab.	pendulum difference (10^{-7} sec.)	mean period 0·50
89	on	54,8	3,9044,5
90	on	50,2	3,9043,9
91	on	56,2	3,9046,0
92	off	55,9	3,9039,2
94	off	53,9	3,9041,6
95	off	63,3	3,9039,0
96	on	54,2	3,9046,0
97	off	58,8	3,9048,2

Mean pendulum periods, stabilized = 0·503,9045,1.
Mean pendulum periods, unstabilized = 0·503,9042,0.
[Pendulums 1 and 3 in use.]

The use of the stabilizer to reduce the middle pendulum to rest when the outer two pendulums were deflected at the beginning of an observation made starting very simple. This process usually requires skill and patience because the two outer pendulums must be released at the instant when the apparatus is level, and if the middle pendulum is moving, this instant is difficult to decide. In an unstabilized run the pendulums were started and then the power supply to the preamplifier was cut off; the magnet current was left on as switching it off would have imparted a jerk to the apparatus.

Table 4 gives a comparison of all stabilized and unstabilized runs made in the submarine. The procedure was to observe for 20 min. at a time, with the stabilizer first on and then off, all other conditions remaining as far as possible constant. Twenty minutes is an adequate time of observation when the depth recorder is available from which to correct for first-order vertical accelerations. The average difference of stabilized from unstabilized observations in table 4 is obviously not significant. The greater standard deviation of the differences between the pendulums for unstabilized swings is a clear indication that sway can occur, and that this sway is largely removed by the stabilizer.

Table 4. Comparison of stabilized and unstabilized pendulum observations

obs. no.	stn. no.	depth dived (ft.)	2nd O.C. (mgal.)	pendulum difference (10^{-7} sec.)	g unst. (980+)	g stab. (980+)	$g_s - g_u$ (mgal.)
1	Portland	—	—	−6,4	1·125	—	−1·2 mean
2	„	—	—	68,2	—	(1·116)	
3	„	—	—	5,1	1·120	—	
4	„	—	—	60,0	—	1·122	
5	„	—	—	−29,2	1·123	—	
6	„	—	—	51,9	—	1·122	
16	„	—	—	60,6	—	1·121	
26	57*b*	90	0·5	49,3	—	1·066	+1
27	57*c*	90	0·5	31,6	1·065	—	—
30	60*a*	180	0·0	58,1	—	1·056	−3
31	60*b*	180	0·0	44,6	1·059	—	—
32	60*c*	65	0·8	56,5	—	1·057	−5
33	60*d*	65	0·8	109,7	1·062	—	—
35	62*a*	90	0·0	43,7	—	1·089	0
36	62*b*	90	0·0	89,3	1·089	—	—
41	66*b*	35	0·8	56,6	—	1·099	−2
42	66*c*	35	1·0	61,7	1·101	—	—
49	69*b*	35	1·4	60,3	—	1·103	−2
50	69*c*	35	1·4	37,3	1·105	—	—
85	52*j*	33	0·5	49,1	—	1·101	+1
86	52*k*	33	0·5	69,3	1·100	—	—
44	Gosport	—	—	60,1	—	1·121	+0·5 mean
45	„	—	—	87,5	1·121	—	
46	„	—	—	56,9	—	1·122	
72	„	—	—	55,1	—	1·122	−4
73	„	—	—	77,5	1·126	—	—
						mean	−1·2
						m.s.e.	±1·8

Mean pendulum differences, stabilized = 55,2 ± 5·1 m.s.e.
Mean pendulum differences, unstabilized = 48,1 ± 40 m.s.e.
2nd O.C. = second-order correction (estimated).
m.s.e. = mean square error.

Although the average difference between the stabilized and unstabilized swings is small, it must be admitted that two errors, of 4 and 5 mgal., have been found. This may have been due to there having been a discrepancy in amplitude at the start, which would give a motion throughout to the centre pendulum and to the short-period pendulum. There would thus be an arc and deviation from isochronism correction which has been neglected as well as possible second-order effects. Provided that closely isochronous pendulums are used and a good antiphase start is made the stabilizer appears to be a considerable advantage on all counts.

Two attempts were made to take an observation on the surface, advantage being taken of the exceptionally calm weather. These failed because of the sliding of the knife-edges on the agates, although the stabilizer was successful in holding the centre pendulum at rest under these conditions. This demonstrates some of the considerable practical difficulties which would beset anyone endeavouring to make pendulum observations in a surface ship. The vibration of the engines, together with sideways accelerations, would send the knife-edges skating along the agates, and lack of perfection in the gimbals would put on such tilts in the direction perpendicular to the swinging-plane that the strain on the knife-edges would be unbearable.

THE AUTOMATIC DEPTH RECORDER

6. Design of the depth Recorder

The instrument is illustrated in figure 5; it was designed by Mr B. C. Browne and manufactured by the Cambridge Instrument Co.

It depends for its action on the measurement by means of a sensitive aneroid bellows of the difference in pressure between the outside of the submarine and a constant pressure produced inside from the compressed air supply. This constant pressure is arranged to be closely equal to the pressure expected to be encountered at the mean depth; the movement of the differential bellows then measures the fluctuations about this mean and a sufficient sensitivity can be obtained over this limited range to give measurements correct to about 1 ft. The movement of the bellows is recorded electrically on an Esterline Angus milliammeter.

Clearly the working substance on one side of the bellows must be easily compressible in order to obtain a mechanical movement of reasonable amplitude. The constant pressure is produced from an air supply and since sea-water would corrode the bellows it is best to have air also on the other side. Hence a device is incorporated for transmitting the sea-pressure to the air-space. This is the diaphragm unit in figure 5.

A strong pressure-tight casting is divided in two by a weak elastic rubber diaphragm; the sea has access to one side and air to the other. A light rod is attached at its centre and projects into a Perspex tube on one side so that the position of the diaphragm may be ascertained by inspection. Two contacts, P, Q are arranged so that when the diaphragm is in its mid-position contact is made to the rod, which is at earth potential. Originally these contacts operated lamps enabling the position of the diaphragm to be ascertained without inspection since the unit had to be placed in an inconvenient position close to the top of the pressure-hull of the submarine. Provided that the volume of air in the pipe between the diaphragm unit and the bellows unit and inside the bellows is small compared with the volume inside the diaphragm unit, and that the elastic membrane is weak, the sea-pressure is communicated unchanged to the bellows over quite a large range of pressure. It is advisable to keep the pipe communicating with the sea short and of large diameter since considerable flow of water may occur in it if the change of depth is rapid.

The bellows unit was provided with a solenoid valve V controlled by a switch to equalize the pressure on either side of the bellows. There was also an exhaust valve

C for releasing the internal pressure when not in use. The gauge 2 was calibrated directly in feet; when the valves *C* and *D* were closed the pressure inside the bellows unit did not vary appreciably with the bellows movement. Auxiliary gauges 1 and 3 enabled the sea-pressure and supply pressure to be read if required.

In operation the valves *V* and *C* were initially open and all others closed. Upon opening *A* and *B* the sea-pressure appeared on gauge 1 and the diaphragm went fully up. *C* was then closed, *V* still remaining open, and air carefully admitted to the system via *F*, *E* and *D* until gauge 2 read the nominal depth of dive, i.e. about the same as gauge 1. At some stage the diaphragm would move hard over to the other side, since this admission of air was not easily controllable. Then by closing *D* and carefully opening *C* the diaphragm could be brought back to centre and *C* and *V* closed. Gauge 2 then read the zero and the sea-pressure brought about movements of the bellows which were recorded.

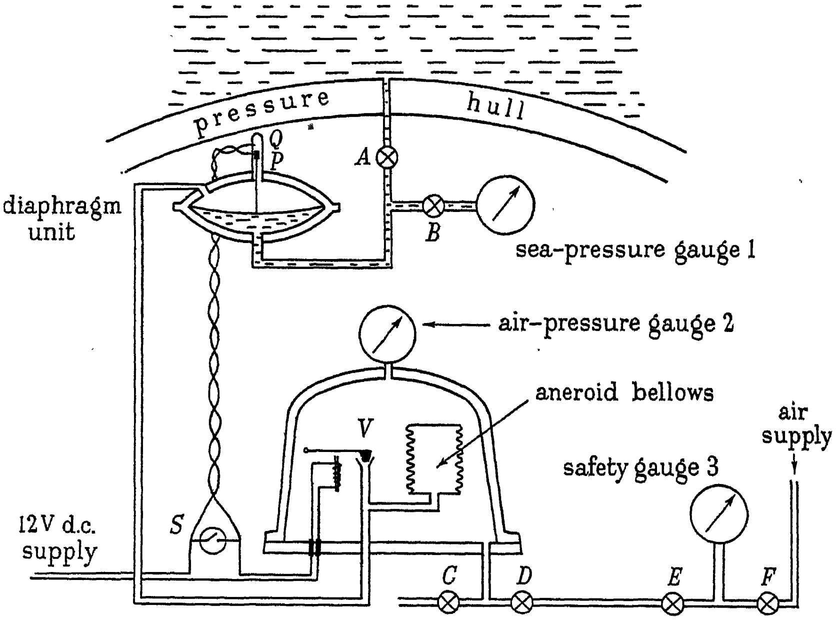


Figure 5. Schematic diagram of depth recorder.

It was found exceedingly difficult to arrange to close *V* as the diaphragm passed centre since it moved so rapidly for small differences of pressure when connected to the whole air space inside the bellows unit. Therefore the contact *Q* was arranged in parallel with the switch controlling the solenoid; *V* then closed instantly as the diaphragm passed centre and afterwards the switch could be made to close the circuit permanently.

The recorder was calibrated by admitting small 'puffs' of air through *D* whilst the system was in operation, the change in the zero being read from gauge 2 and the corresponding step in the record measured. Gauge 2 itself was checked against the other depth gauges in the submarine. The scale was about 2 mm./ft. and a range of ± 25 ft. could be covered.

7. TREATMENT OF DEPTH RECORDS

The pendulums measure the mean value of vertical acceleration throughout the observation. This is

$$\frac{1}{T}\int_0^T g\,dt = g_0 + (\dot{x}_T - \dot{x}_0)/T. \qquad (7{\cdot}1)$$

where $\dot{x}_0$, $\dot{x}_T$ are the vertical velocities at beginning and end respectively. The measurements are obtained by recording the time of coincidence of the pendulum swings, smoothed from the effect of accelerations of wave-period, with an arbitrary frequency standard. This coincidence method is, however, only a special way of counting the number of swings in the time between the instants at beginning and end of the coincidence interval. Therefore each measurement of this interval—and six are taken by subtracting the times of coincidences at beginning and end of the record—is influenced by the value of $\dot{x}$ at each end. Moreover the mean value of $\dot{x}$ over the first and the last few minutes of the observation may be zero and yet the individual values at the instants of coincidence different. Therefore it has been found necessary to determine the slope of the depth/time curve at each instance of coincidence and to correct on the basis of these measurements.

Timing marks had therefore to be put on the depth record during the beginning and end of each run. Two paper speeds were available, 10 cm./min. and 1 mm./min., and the paper was allowed to run at the slow rate during the middle of each observation. The timing marks were applied each minute in the fast sections by manually displacing the pen.

In general, a record had to be very smooth for it to be certain without measurement that no appreciable correction would be necessary. After some experience, however, it was possible to be certain. Most records gave corrections of up to 1 mgal., and a very few about 2 to 3 mgal.; these latter were not always those which appeared on inspection to be the most unsteady. It should be pointed out that in *Talent* our apparatus was mounted in the fore-ends of the submarine where much of the depth-variation was due to tilting of the boat; the variation at the control-room was probably much less. Fluctuations of pressure due to surface waves appeared on the record but were not equivalent to more than 2 to 3 in. of water and were smoothed out.

The fluctuations in depth had a period of about 2 min. for most hydroplane operators. Using this fact we can argue very roughly what the limits to depth-keeping must be if we are to neglect the correction by inspection. For in an observation of 20 min., by 7·1, $\dot{x}$ must not exceed 0·85 cm./sec. if the error is to be less than 1 mgal. for a single measurement of period, or 2·1 cm./sec. for the mean of six. Thus a 2 min. sinusoidal variation must not have amplitude greater than 40 cm. or ± 2 ft. The importance of keeping a depth record is thus clearly demonstrated.

THE NAVIGATIONAL AIDS

8. THE DECCA NAVIGATOR

This well-known system of position-finding depends essentially on comparing the phase of arrival of three waves transmitted on the same frequency (85 kc./sec.) from three coherent shore-stations. In practice, of course, three different rational multiples of a common frequency are used to enable the waves arriving from the separate transmitters to be distinguished; these are then broken down coherently to the fundamental frequency in order to compare phases. Loci of constant phase-difference are clearly hyperbolae and a pattern of intersecting sets of hyperbolae is produced from a master-station and three slave-stations which enables a fix to be obtained. A rotating dial indicates directly the phase-difference between the master and any slave; there is, nevertheless, the possibility that the difference in actual path-length from the two stations will be uncertain to one or more whole wave-lengths. If, however, the apparatus is first set up in a known position it will then be possible to move to any other given position by counting the revolutions of the dials.

The coverage in the English Channel is very good, the transmitters being sited at Buntingford, Herts (master), E. Hoathly, Sussex (green slave), Worm Leighton, Warwickshire (purple slave) and Stoke Holy Cross, Norfolk (red slave). The accuracy of phase comparison is about 1 % of a cycle or 'lane'; in the Channel, lanes are about 5000 yd. wide and hence positions can be found to about 50 yd.

The uncertainty in the whole number of wave-lengths or 'lane-number' introduces a difficulty with submarines. The apparatus of course cannot work during submergence and therefore 'lane-slip' may occur on surfacing out of sight of land. A system of lane-identification has been evolved by Decca but was not available in this instance. However, with the aid of the taut-wire gear, which measured the distance travelled during a dive with high accuracy, and given the approximate direction there was generally little doubt about the number of the lane in which surfacing took place. This could, of course, be checked on returning from the last surfacing position to port. No mistakes ever arose.

9. THE TAUT-WIRE GEAR

By means of a sluice-valve and a special adaptor it is possible to attach a sinker of cast-iron, weighing about 18 lb. to a supply of piano-wire wound on a drum and pass it through the bottom of the hull of the submarine whilst submerged. The sinker lodges in the bottom, and from the rate at which the wire is pulled from the drum the speed of the boat over the ground can be found exceedingly accurately. Also the exact distance run whilst submerged is known.

The wire is of 0·75 mm. gauge and is carried on drums of 9 miles capacity. It passes over a measuring wheel fitted with a pointer and revolution counter, and a tensioning pulley before passing through the grease-packed bush in the adaptor. The revolution counter was read every minute; in this way if the sinker suffers a temporary dislodgement and then catches again the displacement can be checked and no error in speed or distance arises.

Some difficulty was experienced at first due to a tendency for the wire to snap where it was clamped to the sinker; this was overcome by connecting it to the latter by means of a few links of chain which gave a flexible join and allowed the sinker to bounce along the sea-bottom before catching without straining the join. After this precaution had been taken, and provided also that the sinker was released whilst travelling at the lowest possible ground-speed, the wire snapped only once subsequently.

It is clear that this method can only be applied in water whose depth is small compared with the distance travelled during an observation. Errors arise not only on account of the inclination of the wire to the bottom but on account of its tendency to bow out in a tidal stream; these effects have a small influence on the measured speed, however, and the distance is only relevant in the present instance as a guide to lane-identification. If the error in distance is of order 30 yd. in 20 min. the error in speed is about 1/20 knot, equivalent to about 0·2 mgal. if in the east to west direction in these latitudes.

At some double-stations variations in ground-speed due to changes of tidal stream with depth and with time were detected. These variations were of about 0·5 knot east to west and so, if undetected, would have led to errors of about 2½ mgal.

THE SECOND-ORDER CORRECTION

10. Laboratory tests

Since no evidence was obtained during the cruise about the usefulness of applying the second-order correction, laboratory tests were conducted afterwards. The pendulum apparatus, complete, was mounted on a platform suspended by four wires about 20 ft. long from two beams placed across the stair-well in the Department of Geography. A horizontal acceleration of 5 sec. period in a direction either parallel or perpendicular to the swinging plane could thus be applied. The amplitude had to be maintained constant by hand and great care was exercised not to jerk the apparatus during the run, and to keep the accelerations free from components of frequency near the pendulum frequency. This was not easy and under unstabilized conditions large amplitudes were gained by the centre pendulum—not, however, such as to prohibit the experiment.

Table 5. Results of artificial acceleration tests

	obs. no.	ampl. (cm./sec.^{2})	dirn.	2nd O.C. (theor.) (mgal.)	2nd O.C. (meas.) (mgal.)	pendulum difference (10^{-7} sec.)
A	S 14	0	—	0	0	84,9 U
	S 16	7·8	S	15·2	15·5	97,7 St
	S 17	0	—	0	0	83,9 St
	S 18	7·8	P	15·2	17·5	51,3 St
	S 19	7·8	P	15·2	15·0	140,7 U
	S 20	0	—	0	—	—
	S 21	7·8	S	15·2	16·5	80,7 U
B	S 22	9·0	P	20·0	−1	104,3 St
	S 23	5·8	P	9·0	−14	59,4 St
	S 24	0	—	0	0	95,4 St
	S 25	0	—	0	0	92,7 St

A = period 5·0 sec.
B = period 3·1 sec.
S = direction of acceleration parallel to swinging-plane.
P = direction of acceleration perpendicular to swinging-plane.
St = stabilized.
U = unstabilized.

The acceleration was computed from the period and amplitude and the second-order correction calculated from this. Table 5 shows the results and it will be seen that good agreement with theory was obtained both with and without the stabilizer in operation. [The vertical component of acceleration due to the swinging of the apparatus in an arc can easily be seen to be negligible.]

The wires were subsequently shortened to 8 ft. and a period of 3·1 sec. obtained; it will be seen from table 5 that the results under these conditions were anomalous. Unstabilized observations appeared to be impossible, presumably because the motion could not be maintained sinusoidal enough with these crude arrangements and much energy near 1 sec. period reached the pendulums. Insufficient time was available to pursue these investigations as far as would have been desirable.

The conclusion seems to be that the Browne theory applies for accelerations at least down to 5 sec. period; it may apply at 3 sec. but during the latter experiments, the energy spectrum probably extended below this period and it is not possible to say precisely where the limit of inapplicability of the Browne theory comes.

11. CONCLUSION

In the absence of the refinements made it is apparent that large errors might have remained undetected. Putting these together we have: 2½ mgal. possible error due to speed variations; 1½ mgal. due to faulty position finding; and 1½ mgal. due to neglect of the second-order correction. Each of these errors alone is not significant in comparison with the standard error of the pendulum observation itself (about 2½ mgal.) but taken together they will be found to increase the standard error of a gravity measurement to 5 mgal., and consequently some of the larger discrepancies found previously become less perplexing. In general, of course, errors will be nothing like so bad as suggested and the round figure of 3 mgal. usually adopted as the standard error for this kind of work may stand. It is abundantly clear that pendulum observations are unlikely to yield results of greater precision than these. Nevertheless, the importance is demonstrated of keeping careful check on these sources of error. Summarizing our recommendations we may say:

(1) It is probably necessary to continue to carry an accelerometer in order to know when the submarine's motions are small enough for the second-order correction to be neglected;

(2) A depth recorder is essential, if the apparatus cannot be mounted in the control room;

(3) Some form of navigational aid with an accuracy at least as good as 500 yd. is desirable;

(4) In shallow water, taut-wire gear is a substantial advantage.

Development of an automatic recorder for the rate of the fictitious pendulums is a possibility which should be considered now that a stabilizer has been shown to be advantageous. In any case the accuracy of recording could be profitably increased to enable shorter observations to suffice, now that the error arising from depth variation can be corrected by the depth recorder. The value of gravity might then be computed during a voyage which would enable the observing programme to be adjusted to meet the requirements of the geological situation. Considerable saving of labour and expense would result.

The greatest possible thanks are due to the Royal Society and to the Royal Navy, to the former for providing the funds for the expedition and to the latter for making *Talent* available and for the excellent co-operation of all officers.

The author also wishes to thank Lt. I. G. Raikes, captain of *Talent*, and all officers and crew for their willing assistance. Lt. P. G. Satow is to be congratulated on the accuracy and skill of his navigation.

Mr B. C. Browne was in charge of the expedition and proved once more a capable and inspiring leader under arduous circumstances.

Finally, the author wishes to thank the General Board of the University for granting him leave of absence to enable him to join the expedition.

References

Browne, B. C. 1937 *Mon. Not. R. Astr. Soc. Geophys. Suppl.* **4**, **3**, 271–279.

Browne, B. C. & Cooper, R. I. B. 1949 'The British Submarine Gravity surveys of 1938 and 1946.' *Phil. Trans.* A. (in the Press).

Ewing, M. 1937 *Trans. Amer. Geophys. Un.* **18**, Part I, pp. 66–69.

Meinesz, F. A. V. 1929 The theory and practice of pendulum observations at sea. Part I. *Publ. Neth. Geod. Comm.* Delft 1929.

Meinesz, F. A. V. 1937 *Proc. K. Acad. Sci. Amst.* **40**, 650.

Meinesz, F. A. V. 1938 *Proc. K. Acad. Sci. Amst.* **41**, 540.

The passage of a perfect fluid through a critical cross-section or 'throat'

By A. M. Binnie

(*Communicated by Sir Geoffrey Taylor, F.R.S.—Received* 17 *December* 1948)

A review is given of the methods previously adopted of calculating the discharge of a gas through a convergent-divergent nozzle and of a liquid over a broad-crested weir, through a Venturi flume with a horizontal bottom, and in a swirling state through a vertical nozzle. The assumption is made throughout that conditions are uniform over each cross-section of the constriction. Hugoniot's method is shown to be preferable, and it is used to examine the discharge of a liquid through a Venturi flume of any shape and of a swirling gas through a nozzle. It is pointed out that, if an external force is operating in the axial direction, the local velocity of sound or of a long wave, as the case may be, is not attained by the fluid at the exact geometrical throat of the constriction. For the general case of liquid flow, whether swirling or not, the axial velocity of streaming in the absence of external forces is shown to be equal at the throat to the velocity of a long wave.

1. Introduction

The problem of steady fluid motion through a constriction possesses unusual features. The fluid, if a gas, is assumed to fill the channel completely; if a liquid, to possess a free surface. It is supposed that the state of the fluid is specified in a large supply reservoir or at some cross-section upstream from the constriction, and that conditions below the constriction are such that small alterations in them do not affect the flow upstream. Experiment reveals that the flow is then completely determined, and the question arises of calculating the throat conditions and the discharge.

This type of motion appears in a variety of forms; the most important are set out below in the order (as far as can be discovered) in which they have been first considered by previous writers. The rate of change of the cross-section of the stream with respect to distance measured in the direction of streaming is taken as small, hence it is sufficient to regard the main streaming motion as one-dimensional. It is also assumed that the flow is adiabatic and irrotational, thus in cases of swirling motion it is necessary only to consider the distribution of tangential velocity v given by $v = c/r$, where r is the radius and c is a constant. Similar methods apply to the flows of gases and liquids, for an analogy mentioned by Lamb (1932) and by Riabouchinsky (1935) exists between them; a proof was given by Binnie & Hooker (1937*a*).

(i) *Gas flow through a nozzle*

The first successful discussion of this case was by Reynolds (1886*a*, *b*) who derived an expression connecting the pressure p and the mass flow M over a cross-section of area A. By putting $dA/dp = 0$, i.e. making A a minimum, he determined the throat pressure and thus the throat velocity, adding the important observation that the latter is equal to the local velocity of sound. In the method which is stated by Stodola & Lowenstein (1945) to have been originated by Hugoniot (1886), Bernoulli's equation and the equation of continuity of mass-flow M are differentiated with respect to the distance z measured in the general direction of streaming. After dM/dz has been put equal to zero, the differential of the density is eliminated from the two equations, and inspection of the final result shows that the throat velocity and the local velocity of sound are equal. This method was also used by Lamb (1932) and by Taylor (1935). Ewing (1926, 1936), in his two text-books, after employing the first part of Reynolds' method, completed the analysis with the aid of the relation $d(M/A)/dp = 0$ at the throat, based on the consideration that M/A is a maximum there.

(ii) *Liquid flow over a broad-crested weir of uniform breadth*

This type was examined by Unwin (1907), who from Bernoulli's equation derived an expression for the mass-flow in terms of the depth h of the stream over the weir. On putting $dM/dh = 0$ he obtained the critical depth over the weir, but both he and most later writers failed to remark that the corresponding velocity is equal to that of a long wave of small amplitude moving in the stream over the weir.

(iii) *Liquid flow through a Venturi flume having a horizontal bottom*

The method adopted by Unwin was employed by Jameson (1925) to determine the throat velocity for four different shapes of cross-section, but again the point was overlooked that this velocity and that of a long wave at the throat are identical. Hugoniot's method was used by Binnie & Hooker (1937*a*) in their detailed analysis of the rectangular flume.

(iv) *Swirling liquid flow under gravity through a vertical nozzle*

Here the source of supply is an open reservoir, which discharges through a nozzle fixed with its axis vertical, the swirl being sufficiently vigorous to produce an air-core in the constriction. The motion was examined by Binnie & Hookings (1948), who derived an expression for the mass flow involving the radius b of the air-core. To obtain the throat velocity they put $dM/db = 0$ and showed that this velocity is equal to that of an axial wave moving in the spinning liquid at the throat.

2. A study of these references reveals that those writers who have employed the mass-flow criterion have given little justification for their analysis beyond demonstrating that the throat velocity so calculated is equal to the local velocity of sound or of a long wave, as the case may be. Even this measure of support was omitted in cases (ii) and (iii). In its usual form the argument goes on to state that the velocity in the divergent part of the stream exceeds the local velocity of sound, hence disturbances near the exit cannot travel upstream and alter the discharge. This appeal to physical reasoning is certainly plausible but perhaps not wholly convincing. It should be remarked that the disturbances considered must be small. Clearly it is possible to stop the flow by closing a valve or sluice at the exit, thus causing a wave of finite amplitude to travel up the constriction.

Hugoniot, who confined his attention to gases moving in one dimension under no external forces, proved that the throat velocity must necessarily be equal to the local velocity of sound. His method can be readily extended to include liquids, external forces and swirling motions, and in the following pages it is used to examine the flow under gravity of a liquid through a Venturi flume of any shape and through a vertical nozzle. It will be shown that of the methods which are available his is to be preferred, because with no greater labour it yields more information than the others. A general proof is also given of the equality of the throat velocity in liquid flow with the local velocity of a long wave. Lastly, the same method is employed to solve the problem of swirling gas flow through a nozzle.

3. LIQUID FLOW THROUGH A VENTURI FLUME

As a first example of Hugoniot's method we will consider the general form of Venturi flume having a humped bottom and walls not necessarily vertical. At a typical cross-section, let q be the velocity, h the depth of the liquid and H the height of the bottom above a fixed horizontal datum. Bernoulli's equation for every stream-line is

$$h+H+\frac{q^2}{2g} = h_0, \tag{3·1}$$

where h_0 is the height of the reservoir surface above the datum, hence

$$\frac{dh}{dz}+\frac{dH}{dz}+\frac{q}{g}\frac{dq}{dz} = 0. \tag{3·2}$$

The mass flow M is given by

$$\rho q A = M, \tag{3·3}$$

A denoting the area of the cross-section of the stream and ρ the constant density, therefore, since M is independent of z,

$$q\frac{dA}{dz}+A\frac{dq}{dz}=\frac{1}{\rho}\frac{dM}{dz}=0. \tag{3·4}$$

Now A is a function of h and z, and we have that

$$\frac{\partial A}{\partial h}=B, \tag{3·5}$$

and

$$\frac{\partial A}{\partial z}=h\frac{\partial B_1}{\partial z}, \tag{3·6}$$

where B is the surface breadth and B_1 the mean breadth, hence

$$\frac{dA}{dz}=\frac{\partial A}{\partial h}\frac{dh}{dz}+\frac{\partial A}{\partial z}=B\frac{dh}{dz}+h\frac{\partial B_1}{\partial z}. \tag{3·7}$$

On eliminating dA/dz and dh/dz from (3·2), (3·4) and (3·7) we obtain

$$\left(q^2-\frac{gA}{B}\right)\frac{B}{gq}\frac{dq}{dz}+B\frac{dH}{dz}-h\frac{\partial B_1}{\partial z}=0. \tag{3·8}$$

We suppose that the summit of the humped bottom and the minimum mean breadth occur at the same cross-section. Then at the throat $dH/dz=\partial B_1/\partial z=0$, and also $dq/dz \neq 0$, since we are interested in a type of flow asymmetrical with respect to the throat cross-section. Hence (3·8) shows that the throat velocity is given by

$$q^2=\frac{gA}{B}, \tag{3·9}$$

and we recognize that this velocity is identical with that of a long wave at the throat. Expressions for the throat depth and for M can then be derived from (3·1) and (3·3).

A further point is made evident by (3·8). In the convergent part of the constriction dH/dz is positive and $\partial B_1/\partial z$ negative, hence if dq/dz is to be positive, as we wish, q^2 must be less than gA/B. To produce the required type of flow, the velocity of the stream entering the constriction must be less than the local velocity of a long wave. When this condition is violated, the depth of the liquid increases and the velocity falls as the throat is approached.

The alternative procedure of putting $dM/dh=0$ at the throat may be followed with the aid of (3·4), (3·7) and (3·2). Used together they yield

$$\frac{1}{\rho}\frac{dM}{dh}=\frac{B}{q}\left(q^2-\frac{gA}{B}\right) \tag{3·10}$$

at the throat. To take $dM/dh=0$ at the throat is equivalent to supposing that $q^2=gA/B$ there; consequently, for this method to be logical, the initial assumption requires justification. We assert that $dM/dz=0$ at every cross-section and that $dh/dz \neq 0$ at the throat. Hence, on differentiating with respect to z the expression for M in terms of h and A derived from (3·1) and (3·3), we may, after employing (3·7), take the coefficient of dh/dz as zero at the throat for there the other terms disappear.

This is, of course, a process identical with putting $dM/dh = 0$ at that cross-section. But the method, it will be noticed, provides no information of the kind described in the preceding paragraph.

4. SWIRLING LIQUID FLOW THROUGH A NOZZLE

Hugoniot's method will next be applied to a swirling liquid flowing under gravity through a vertical nozzle (case iv). It was shown by Binnie & Hookings (1948) that, for a horizontal cross-section in the nozzle at a distance z below the free surface in the reservoir, Bernoulli's equation takes the form

$$2gz = w^2 + \frac{c^2}{b^2}, \tag{4·1}$$

where c is the irrotational constant, b the radius of the central core of the vortex, and w the vertical velocity which is uniform over the cross-section. The mass flow across the section is

$$\rho w \pi (a^2 - b^2) = M, \tag{4·2}$$

a denoting the radius of the nozzle. On differentiating (4·1) and (4·2) we obtain

$$g = w\frac{dw}{dz} - \frac{c^2}{b^3}\frac{db}{dz} \tag{4·3}$$

and

$$(a^2 - b^2)\frac{dw}{dz} + 2w\left(a\frac{da}{dz} - b\frac{db}{dz}\right) = \frac{1}{\rho\pi}\frac{dM}{dz} = 0. \tag{4·4}$$

From these db/dz can be eliminated, the resulting equation being

$$g = (w^2 - V^2)\frac{1}{w}\frac{dw}{dz} - \frac{c^2 a}{b^4}\frac{da}{dz}, \tag{4·5}$$

where (as shown by Binnie & Hookings)

$$V = \left\{\frac{c^2(a^2 - b^2)}{2b^4}\right\}^{\frac{1}{2}} \tag{4·6}$$

is the axial velocity of a long wave in the spinning liquid. From (4·5) it is seen that this velocity and the streaming velocity are equal, not at the geometrical throat where $da/dz = 0$, but at a cross-section above it defined by

$$\frac{da}{dz} = -\frac{gb^4}{c^2 a}. \tag{4·7}$$

On putting $w = V$ in (4·3) and using (4·6) and (4·7) we find that the corresponding inclination of the inner surface is

$$\frac{db}{dz} = -\frac{3gb^5}{c^2(2a^2 + b^2)}. \tag{4·8}$$

Evidently the last two equations arise from the existence of an external force acting in the direction of streaming, and, if their right-hand sides are of appreciable magnitude, they are at variance with the basic assumption, explained in § 1, that conditions over every cross-section are approximately uniform. This point is missed if reliance is placed solely on the condition $dM/db = 0$ at the throat, which yields the correct throat velocity but no further information. That this condition is here not strictly accurate may be seen by adopting the procedure outlined at the end of § 3.

It appears that these effects were small in the experiments described by Binnie & Hookings. With the use of (4·7) and (4·8), calculations show that, in the second set of observations given in their table 1, $-da/dz$ varied from 0·046 at the lowest head to 0·024 at the highest, the corresponding values of $-db/dz$ ranging from 0·027 to 0·018. Thus the local velocity of a long wave was theoretically attained at a cross-section very close to the geometrical throat.

5. GENERAL CASE OF LIQUID FLOW

We now proceed to consider why the throat velocity must necessarily be always identical with the local velocity of a long wave, and as a first step the general expression for the latter will be derived.

As generalized co-ordinates we take z in the direction of the streaming motion and α at right angles to it. The cross-section of the liquid is assumed to be uniform and the z-axis lies outside the liquid, the ordinates of both the free and the fixed surfaces being positive and the latter exceeding the former. Let the pressure distribution in the liquid be given by

$$\frac{p}{\rho} = F(\alpha), \tag{5·1}$$

so that, if the free surface where the pressure is zero is situated at $\alpha = b$, the absolute pressure at the point α, z in the liquid is

$$\frac{p}{\rho} = F(\alpha) - F(b), \tag{5·2}$$

When an infinitely small wave of surface elevation η exists at the cross-section under consideration, the pressure at α, z becomes

$$\frac{p}{\rho} = F(\alpha) - F(b-\eta), \tag{5·3}$$

and then

$$\frac{1}{\rho}\frac{\partial p}{\partial z} = F'(b)\frac{\partial \eta}{\partial z}. \tag{5·4}$$

Provided there are no external forces acting in the z direction the equation of motion is

$$\frac{\partial w}{\partial t} = -\frac{1}{\rho}\frac{\partial p}{\partial z}, \tag{5·5}$$

the velocity being denoted by w. Taking ξ to be the displacement, we find from (5·4) and (5·5) that

$$\frac{\partial^2 \xi}{\partial t^2} = \frac{\partial w}{\partial t} = -F'(b)\frac{\partial \eta}{\partial z}. \tag{5·6}$$

Now it is shown in text-books that the equation of continuity is

$$A\frac{\partial \xi}{\partial z} + B\eta = 0, \tag{5·7}$$

where A is the cross-section and B the breadth of the surface. The result of eliminating ξ from (5·6) and (5·7) is

$$\frac{\partial^2\eta}{\partial t^2} = \frac{A}{B}F'(b)\frac{\partial^2\eta}{\partial z^2}, \tag{5·8}$$

which represents a wave moving with velocity V given by

$$V^2 = \frac{A}{B}F'(b). \tag{5·9}$$

We now turn to the flow through a constriction. Since the pressure is given by (5·2) the general form of Bernoulli's equation is

$$\frac{1}{2}\left(w^2+\frac{c^2}{\alpha^2}\right)+\Omega+F(\alpha)-F(b) = \text{constant}, \tag{5·10}$$

c being the irrotational constant. We restrict ourselves to the case where the potential Ω of the external forces is such that $\partial\Omega/\partial z = 0$. Hence, for a stream-line along the fixed surface $\alpha = a$,

$$\frac{1}{2}\left(w^2+\frac{c^2}{a^2}\right)+\Omega+F(a)-F(b) = \text{constant}, \tag{5·11}$$

and on differentiation we have

$$w\frac{dw}{dz}-\frac{c^2}{a^3}\frac{da}{dz}+F'(a)\frac{da}{dz}-F'(b)\frac{db}{dz} = 0. \tag{5·12}$$

The mass flow is given by

$$wA = \frac{M}{\rho}, \tag{5·13}$$

hence

$$w\frac{dA}{dz}+A\frac{dw}{dz} = \frac{1}{\rho}\frac{dM}{dz} = 0. \tag{5·14}$$

Now A is a function of z and b, therefore

$$\frac{dA}{dz} = \frac{\partial A}{\partial b}\frac{db}{dz}+\frac{\partial A}{\partial z} = -B\frac{db}{dz}+A'\frac{da}{dz}, \tag{5·15}$$

where A' is the breadth of the fixed surface. On eliminating db/dz and dA/dz from (5·12), (5·14) and (5·15) we arrive at

$$\frac{1}{w}\left(w^2-\frac{A}{B}F'(b)\right)\frac{dw}{dz}+\left(F'(a)-\frac{c^2}{a^3}-\frac{A'}{B}F'(b)\right)\frac{da}{dz} = 0. \tag{5·16}$$

At the throat $da/dz = 0$ and $dw/dz \neq 0$, hence the velocity there is given by

$$w^2 = \frac{A}{B}F'(b), \tag{5·17}$$

and on comparing this with (5·9) we see that

$$w = V. \tag{5·18}$$

For the broad-crested weir (case ii) and the Venturi flume (case iii) $F(\alpha) = g\alpha$, hence $F'(b) = g$ and $w^2 = gA/B$ in accordance with (3·9). For swirling liquid flow through a nozzle

$$F(\alpha) = \int \frac{v^2}{\alpha}\,d\alpha = \int \frac{c^2}{\alpha^3}\,d\alpha. \tag{5·19}$$

Hence $F'(b) = c^2/b^3$ and $$w^2 = \frac{\pi(a^2-b^2)}{2\pi b}\frac{c^2}{b^3} = \frac{c^2(a^2-b^2)}{2b^4}, \tag{5·20}$$

which agrees with (4·6).

6. Swirling gas flow through a nozzle

Finally we will examine the problem of swirling gas flow through a nozzle in the absence of external forces.

Rayleigh (1916) showed that, for steady flow symmetrical about the axis of z, the general equations of motion are

$$\left.\begin{aligned} u\frac{\partial u}{\partial r} - \frac{v^2}{r} + w\frac{\partial u}{\partial z} &= -\frac{\partial P}{\partial r},\\ u\frac{\partial v}{\partial r} + \frac{uv}{r} + w\frac{\partial v}{\partial z} &= 0,\\ u\frac{\partial w}{\partial r} \qquad + w\frac{\partial w}{\partial z} &= -\frac{\partial P}{\partial z}.\end{aligned}\right\} \tag{6·1}$$

Here $$P = \int \frac{dp}{\rho}, \tag{6·2}$$

and the velocities u, v, w are reckoned in the directions of r, θ, z increasing: it is assumed that there are no external forces. It will be seen from these equations that in spite of compressibility effects irrotational flow is possible, in which vr = constant and w is uniform over each cross-section, provided that u is negligible. But the velocity at the axis would be infinite, therefore the nozzle must be of annular cross-section.

On combining Bernoulli's equation with the isentropic relation p/ρ^γ = constant we obtain

$$\frac{\rho}{\rho_0} = \left(1 - \frac{\gamma-1}{2}\frac{q^2}{q_0^{*2}}\right)^{1/(\gamma-1)} \tag{6·3}$$

where γ is the ratio of the specific heats, ρ is the density of the gas when it is moving with velocity q, and ρ_0 is the density in the reservoir where the local velocity of sound is q_0^*. In the present circumstances $q^2 = w^2 + c^2/r^2$, and the mass flow over a cross-section of outer radius a and inner radius b is

$$M = \int_b^a w\rho_0\left(1 - \frac{\gamma-1}{2}\frac{w^2+c^2/r^2}{q_0^{*2}}\right)^{1/(\gamma-1)} 2\pi r\,dr. \tag{6·4}$$

This expression can be integrated in finite terms if $(\gamma-1)^{-1}$ is an integer or a half-integer.

Although no gas is known for which $\gamma = 2$, it is worth while briefly considering this case, for results can be obtained and trends noted without laborious calculations. With this value of γ (6·4) reduces to

$$\frac{Mq_0^{*2}}{\pi\rho_0} = w\left\{\tfrac{1}{2}(a^2-b^2)(2q_0^{*2}-w^2)-c^2\log\frac{a}{b}\right\}. \tag{6·5}$$

When this expression is differentiated with respect to z and dM/dz put equal to zero, the result is

$$0 = \left\{\tfrac{1}{2}(a^2-b^2)(2q_0^{*2}-3w^2)-c^2\log\frac{a}{b}\right\}\frac{dw}{dz} + \text{terms involving } da/dz \text{ and } db/dz, \tag{6·6}$$

therefore the throat value of w is given by

$$\frac{w^2}{q_0^{*2}} = \frac{2}{3}\left(1-\frac{c^2}{q_0^{*2}a^2}\frac{\log a/b}{1-b^2/a^2}\right). \tag{6·7}$$

When $c = 0$, (6·7) reduces to $w^2/q_0^{*2} = 2/3 = 2/(\gamma+1)$ as it should. When $b \to a$, (6·7) becomes

$$\frac{w^2}{q_0^{*2}} = \frac{2}{3}\left(1-\frac{c^2}{2q_0^{*2}a^2}\right), \tag{6·8}$$

as can also be proved directly by applying the same method to a nozzle in which $(a-b)$ is very small throughout. Now the relation between q and q^*, the local velocity of sound, is shown by Lamb (1932) to be

$$q^2+\frac{2q^{*2}}{\gamma-1} = \frac{2q_0^{*2}}{\gamma-1}, \tag{6·9}$$

which here becomes

$$\frac{w^2+2q^{*2}}{q_0^{*2}} = 2\left(1-\frac{c^2}{2q_0^{*2}a^2}\right). \tag{6·10}$$

On comparing this with (6·8) we see that when $b \to a$ the throat value of w is equal to the local velocity of sound. A similar result was mentioned by Taylor (1930) and by Binnie & Hooker (1937*b*), who pointed out that in the two-dimensional spiral flow of a gas starting from rest at infinity, the radial component of velocity cannot exceed the local velocity of sound. It may easily be shown from (6·3) and (6·7) that at the throat

$$\frac{w^2}{q_0^{*2}} = \frac{1}{\rho_0}\frac{\int_b^a \rho 2\pi r\,dr}{\pi(a^2-b^2)}, \tag{6·11}$$

which is the ratio of the mean density over the throat cross-section to the density in the reservoir. This result is identical with that obtained when c is zero. Equation (6·7) is subject to the restriction that the maximum possible velocity is given by

$$q^2 = \frac{2q_0^{*2}}{\gamma-1}. \tag{6·12}$$

This limit, when applied at $r = b$ where the velocity is greatest, yields the condition

$$\frac{w^2}{q_0^{*2}} \ngtr 2-\frac{c^2}{q_0^{*2}b^2}, \tag{6·13}$$

which is not inherent in (6·7) for that equation, derived from the relation p/ρ^2 = constant, is theoretically applicable to negative as well as to positive densities. Numerical results, expressed in non-dimensional form, are shown in figure 1. The full lines give the throat values of w/q_0^*, obtained from (6·7) and (6·8), and plotted on a base of b/a for $c^2/(q_0^* a)^2 = 0, 0{\cdot}5$, 1 and 1·5: the curve for $c^2/(q_0^* a)^2 = 2$ reduces to the point $w/q_0^* = 0, b/a = 1$, where the velocity is wholly tangential. These lines are terminated on the left by intersection with the curves of maximum possible velocity, calculated from (6·13) and shown chain-dotted. Along the broken line joining the points of intersection the pressure on the inner wall of the throat is zero; flow is not possible under conditions represented by points below this line, and if attempted, the nozzle would not run full.

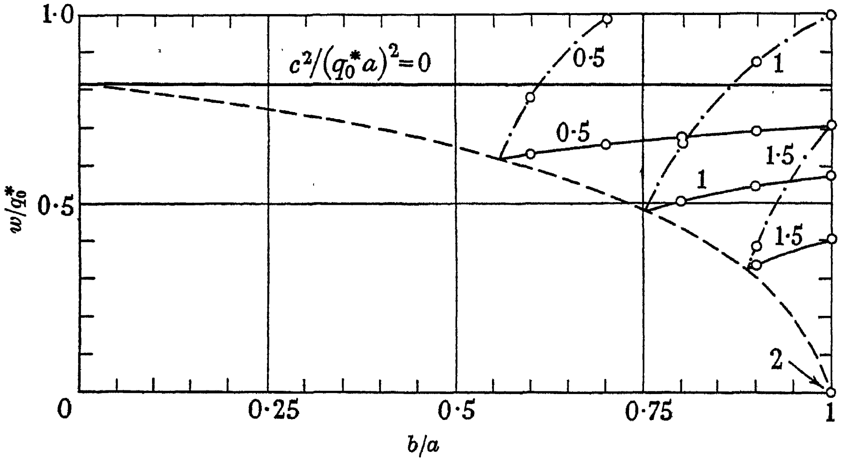


FIGURE 1. Swirling gas flow. Throat velocities when $\gamma = 2$.

Calculations on the same lines were made for the case $\gamma = 1{\cdot}4$, i.e., $(\gamma - 1)^{-1} = \frac{5}{2}$, which is a close approximation to the accepted value for air. The details are too lengthy to be given here, but the resulting equation for the throat velocity, corresponding to (6·7), is

$$\begin{aligned}(1-x-s)^{\frac{1}{2}}\{(1-x+2s)(\tfrac{1}{2}-3x+\tfrac{4}{3}s)-3s^2\}& \\ -\frac{1}{y}(1-x-sy)^{\frac{1}{2}}\{(1-x+2sy)(\tfrac{1}{2}-3x+\tfrac{4}{3}sy)-3s^2y^2\}& \\ +\tfrac{5}{2}s(1-x)^{\frac{1}{2}}(1-4x)\log\left\{y^{\frac{1}{2}}\frac{(1-x)^{\frac{1}{2}}-(1-x-s)^{\frac{1}{2}}}{(1-x)^{\frac{1}{2}}-(1-x-sy)^{\frac{1}{2}}}\right\} &= 0, \qquad (6{\cdot}14)\end{aligned}$$

where

$$x = \frac{w^2}{5q_0^{*2}}, \quad y = \frac{a^2}{b^2} \quad \text{and} \quad s = \frac{c^2}{5q_0^{*2}a^2}. \qquad (6{\cdot}15)$$

After the insertion of specified values of y and s into this expression, x can be found by trial. When $b \to a$,

$$\frac{w^2}{q_0^{*2}} = \frac{5}{6} - \frac{c^2}{6q_0^{*2}a^2}, \qquad (6{\cdot}16)$$

as may be proved either directly or from (6·14). The maximum possible velocity is shown by (6·12) to be given by $q^2 = 5q_0^{*2}$, and on applying this result at $r = b$ we find that $(1-x-sy)$ must be positive or zero. This restriction is seen to be automatically imposed in (6·14). At the limit where

$$1-x-sy = 0, \tag{6·17}$$

(6·14) reduces to

$$(1-x-s)^{\frac{1}{2}}\{(1-x+2s)(\tfrac{1}{2}-3x+\tfrac{4}{3}s)-3s^2\} \quad . \\ +\tfrac{5}{2}s(1-x)^{\frac{1}{2}}(1-4x)\log[s^{-\frac{1}{2}}\{(1-x)^{\frac{1}{2}}-(1-x-s)^{\frac{1}{2}}\}] = 0. \tag{6·18}$$

In (6·18) the value of x can be determined by trial when s is given, and the corresponding value of y is then found from (6·17). Numerical results obtained from (6·16), (6·18) and (6·14) are shown in figure 2, where again the throat values of w/q_0^* are shown by full lines and the limit by a broken line. Apart from a considerable increase in the range of possible values of $c^2/(q_0^*a)^2$ this diagram has much the same form as figure 1.

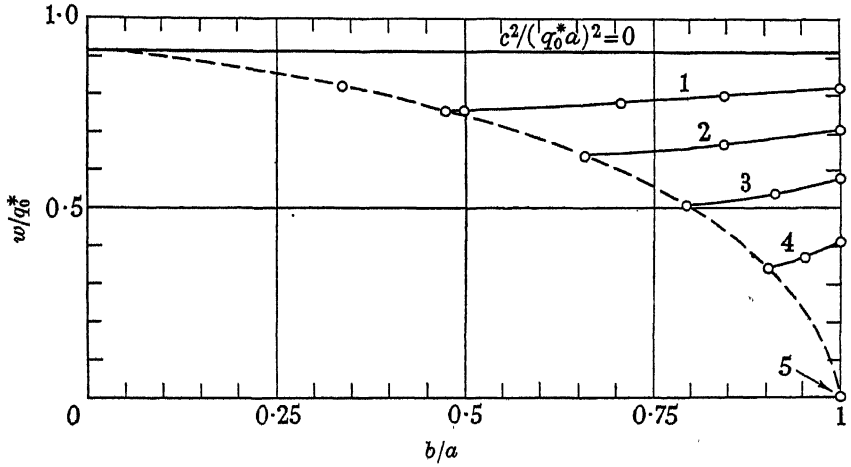


FIGURE 2. Swirling gas flow. Throat velocities when $\gamma = 1·4$.

REFERENCES

Binnie, A. M. & Hooker, S. G. 1937*a* *Proc. Roy. Soc.* A, **159**, 592.
Binnie, A. M. & Hooker, S. G. 1937*b* *Phil. Mag.* (7) **23**, 597.
Binnie, A. M. & Hookings, G. A. 1948 *Proc. Roy. Soc.* A, **194**, 398.
Ewing, Sir J. A. 1926 *The steam engine*, p. 224. Cambridge University Press.
Ewing, Sir J. A. 1936 *Thermodynamics for engineers*, p. 200. Cambridge University Press.
Hugoniot, H. 1886 *C.R. Acad. Sci., Paris*, **103**, 1178.
Jameson, A. H. 1925 *Proc. Inst. Water Engrs*, **30**, 19.
Lamb, Sir H. 1932 *Hydrodynamics*, pp. 26, 477. Cambridge University Press.
Rayleigh, Lord 1916 *Proc. Roy. Soc.* A, **93**, 148 (also *Papers*, **6**, 447).
Reynolds, O. 1886*a* *Proc. Manchr Lit. Phil. Soc.* **25**, 55.
Reynolds, O. 1886*b* *Phil. Mag.* (5) **21**, 185 (also *Papers*, **2**, 311).
Riabouchinsky, D. P. 1935 *J. R. Aero. Soc.* **39**, 437.
Stodola, A. & Loewenstein, L. C. 1945 *Steam and gas turbines*, p. 44. New York: Peter Smith.
Taylor, Sir G. I. 1930 *J. Lond. Math. Soc.* **5**, 224.
Taylor, Sir G. I. 1935 *Aerodynamic theory* (ed. Durand), **3**, 226. Berlin: Springer.
Unwin, W. C. 1907 *Treatise on hydraulics*, p. 107. London: Black.

Residual lattice strains in plastically deformed polycrystalline metal aggregates

By G. B. Greenough, Ph.D.

Cavendish Laboratory, University of Cambridge and Royal Aircraft Establishment, Farnborough

(*Communicated by E. Orowan, F.R.S.—Received* 28 *December* 1948)

The residual lattice strain observed by X-ray diffraction methods in plastically extended polycrystalline aggregates has been investigated. It is shown that, if several diffraction lines are examined, the residual lattice strains vary both in magnitude and in sign.

It is known that one of the causes of internal stresses in a plastically deformed polycrystalline aggregate is the variation with orientation of the yield tensions of the crystals in the aggregate. Since only grains with certain orientations contribute to a given diffraction line, the lattice strain measured is not the mean strain in all the crystals, which should be zero, but the average of the strains in those crystals with particular orientations, which may be positive or negative. It is considered that this constitutes one of the reasons for the hitherto unexplained observations which have been made on residual lattice strains.

Quantitative conclusions from this explanation are in satisfactory agreement with observations.

1. Introduction

The lattice strains occurring in the crystals of a polycrystalline metal specimen when the aggregate is subjected either to externally applied loads or to internal stresses may be determined from observations on the positions of the X-ray lines diffracted by the aggregate. The position of the maximum of the X-ray diffraction line depends on the average of the spacing between many planes in all those crystals which contribute to the diffraction line. Small variations in this spacing about the mean value result in a broadening of the line, but changes in the position of the maximum of the line indicate changes in the average interplanar spacing.

Observations by Bollenrath, Hauk & Osswald (1939), by Smith & Wood (1941), and by others, on the positions of the maxima of diffraction lines have shown that when a specimen has been plastically extended and the applied stress removed, there remains a mean lattice strain in the grains. This mean residual lattice strain may be attributed to a general lattice expansion (Smith & Wood 1941), or to internal stresses. Experimental evidence (Smith & Wood 1944; Bollenrath *et al.* 1939) supports this latter view, and the hypothesis of a general lattice expansion has been discarded.

Since after deformation there are no external forces acting on the specimen, the sum of the stresses across any plane section of the specimen must be zero. Thus, if the mean residual lattice strain in the part of the specimen contributing to the diffraction line is interpreted as due to a mean stress there, some other portion of the specimen must carry a stress of opposite sign. Such a stress system, where stresses in one part of the deformed aggregate are balanced by those in another, would arise in the aggregate if different parts had different yield tensions; then after the plastic extension of the specimen, the 'hard' parts would remain in tension and the 'soft' parts would be driven into compression.

Bollenrath *et al.* (1939) have suggested that the surface of the specimen has a lower yield stress than the interior, and after the plastic extension of the specimen the surface and interior of the specimen are in opposite states of stress. They have described experiments which showed that the mean residual lattice strains in a plastically extended bar, at surfaces successively exposed by etching away part of the metal, decreased as the specimen became thinner. This is contrary to experimental work described by Smith & Wood (1944), who found that the residual lattice strain remained approximately constant at surfaces exposed by etching specimens. Davidenkov & Timofeeva (1946) made mechanical strain measurements on plastically extended alluminium sheets as surface layers were etched away, and showed that the surface was not in a state of uniaxial macroscopic stress. Although earlier work by the author (Greenough 1948) supported the idea that no surface macroscopic stress exists, more recent results indicate that this early experiment was not properly designed and that a macroscopic stress system may explain a part, but not all, of the observed residual lattice strains.

An alternative explanation could be based on that put forward during a discussion in 1943 between Bragg, Orowan, Smith & Wood (unpublished). It is possible that, during the process of plastic deformation, the accumulation of dislocations round obstacles to their propagation through the crystal would cause higher stresses there than in the remaining bulk of the crystal. After the removal of the applied load producing the deformation, stresses would arise in the crystal, those in the neighbourhood of the arrested dislocations being balanced by those in the bulk of the crystal remote from the dislocations. The lattice spacings in the region of the arrested dislocations would be insufficiently regular to contribute to the diffraction line, but the strain in the bulk of the crystal would be sufficiently uniform to produce a movement of the diffraction line. Smith & Wood (1944) put forward a similar hypothesis in which the grain boundaries were postulated as becoming amorphous.

These two explanations assume that the part of the specimen stressed to balance the part with the residual lattice strain is either too far below the surface to be irradiated by the X-rays, or insufficiently crystalline to contribute to the diffraction.

A third explanation could be based on the fact that only grains with certain orientations in the irradiated area of the aggregate contribute to the diffracted beam, a point illustrated by figure 1. The lattice strain determined from measurements on X-ray diffractions is a mean of the strains in those grains whose 'reflecting' normals are in fixed directions, and this may differ from the mean taken over all the grains in the specimen. The residual lattice strain could be due to stresses in the reflecting grains which are balanced by stresses of opposite sign in neighbouring grains, whose orientations are such that they do not contribute to the reflexion.

The explanation of the development of residual lattice strains can be made explicit by discussing the particular case of a polycrystalline specimen extended by an applied tension. The tensile stress required to cause glide in a given grain, the 'yield tension' of the grain, depends on the orientation of the glide planes and glide directions. During the plastic extension of the aggregate, grains with a higher yield tension carry a greater stress, and hence contain a greater elastic strain, than their neighbours, whose orientations are such that they glide more easily. When the applied stress is removed, grains with the greater elastic strain compress their neighbours, so that grains with a high yield tension will be left in tension and the softer ones in compression.

Among the grains that, by virtue of their special orientation, give rise to a given reflexion line, there will be, say, more of a 'soft' than of a 'hard' orientation. Hence the average residual stress in these crystals will be compressive and the average spacing between the reflecting planes, measured perpendicular to the direction of elongation of the aggregate, will increase. A second diffraction line which receives reflexions from grains of a different orientation which is 'hard' will show a residual lattice strain of opposite sign. Thus, the residual lattice strains determined by the examination of many diffraction lines should have values which vary in sign and in magnitude; since the mean stress in any plane across the specimen is zero, the appropriate mean value of all the residual lattice strains should also be zero.

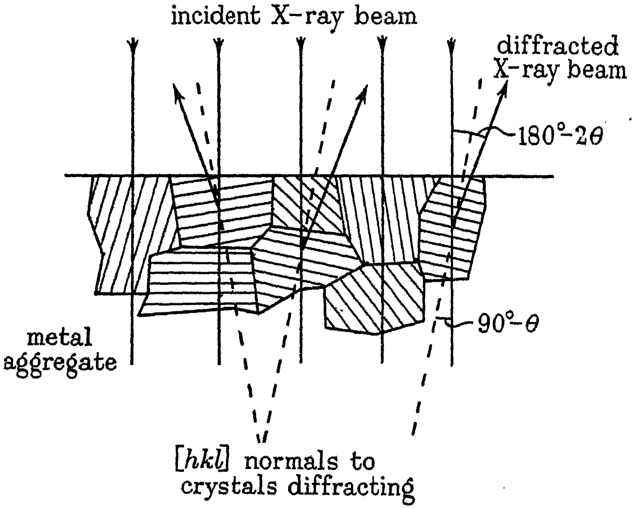


FIGURE 1. Crystals contributing to *hkl* diffraction line.

2. EXPERIMENTAL DETERMINATION OF RESIDUAL LATTICE STRAINS

Experimental work was undertaken to measure the lattice strains remaining in various metal specimens after they had been plastically extended and unloaded. No attempt was made to construct stress against lattice-strain curves of the type described by Smith & Wood by observing the lattice strains when the specimens were subjected to loads.

(a) *Method of measuring lattice strains*

To simplify the determination of accurate interplanar spacings, wire specimens were used and a Bradley-Jay camera of 19 cm. diameter employed to record the diffraction lines. It was shown experimentally that, to determine accurately changes in interplanar spacing for a given metal, it was not necessary to correct for camera errors, other than film shrinkage, for all diffraction lines for which the Bragg angle θ was greater than 65°, provided that the specimens were of the same diameter.

If the Bragg angle of diffraction for the deformed specimen is $\theta+\delta\theta$, where θ is the corresponding angle for the annealed material, then the lattice strain $\delta d/d$ is given by the equation

$$\delta d/d = -\cot\theta\delta\theta,$$

where d is the average spacing between the 'reflecting' planes.

The volume of specimen irradiated by X-rays was large enough to give diffraction lines of uniform density when the specimen was rotated about its own axis. This ensured that the behaviour of the line represented that of the average grain of the orientation reflecting, and that random deviations from this average due to variations in the orientations of the neighbouring grains were eliminated.

(*b*) *Examination of specimens*

The metals examined were nickel, copper and aluminium, all of commercial purity. Drawn wires of 0·5 mm. diameter were annealed at a temperature above that at which recrystallization occurred until the wires were completely free from strain. Then, since it has been shown by Smith & Wood (1944), and checked in these experiments, that annealing specimens containing residual lattice strains removes all the strains, the following experimental procedure was employed.

A specimen of the wire was extended and the plastic extension in the uniform region determined from the reduction of area. A diffraction photograph was taken of the specimen in the region of uniform extension; this gave the value of the lattice spacing in the strained metal. The specimen was then annealed and a further photograph taken to obtain the strain-free spacing. The residual lattice strains in the plastically extended metal were determined for those reflexions for which $\theta > 65°$. Several X-radiations were used so that various reflexion lines could be investigated for each metal; in each case five different specimens were examined and a mean of the separate results used to reduce random errors arising from the measurement of broadened diffraction lines.

(*c*) *Results*

With the arrangement of specimen and X-ray camera used, the axis of the wire was perpendicular to the plane containing the incident and diffracted X-ray beams. Thus, the strains measured, which are in the directions of the reflecting normals, were always in a direction perpendicular to the direction of the applied tension producing the plastic deformation. A residual lattice strain that corresponds to an increase of the interplanar spacing after the plastic extension of the specimen is regarded as positive.

Some reflexion lines were recorded with more than one X-radiation, and in all cases the results from the five separate specimens and from the different X-radiations agreed with each other within the limits of experimental error. Table 1 gives the mean residual lattice strains observed for the various planes and the X-radiations used. The probable value of the random part of the error in the mean residual lattice strain was estimated from the individual results; it lay between $\pm 1\times 10^{-5}$ and $\pm 3\times 10^{-5}$.

These results are very similar, in general, to those given previously (Greenough 1947) for magnesium and iron.

TABLE 1. RESIDUAL LATTICE STRAINS IN PLASTICALLY EXTENDED WIRES

(*a*) Aluminium. Plastic extension 15 %

plane reflecting	420	331	333 511	422	400	222	311
X-radiation used	Fe $K\beta$ Co $K\alpha$	Fe $K\beta$ Co $K\alpha$	Cu $K\alpha$	Co $K\beta$ Cu $K\alpha$	Fe $K\alpha$	Cr $K\alpha$	Cr $K\alpha$
residual lattice strain ($\times 10^{-5}$)	+5	$-0{\cdot}_5$	+7	+5	+8	−3	+8

(*b*) Copper. Plastic extension 16 %

plane reflecting	400	422	331	311	222
X-radiation used	Co $K\alpha$	Zn $K\alpha$	Co $K\alpha$	Mn $K\alpha$	Mn $K\beta$
residual lattice strain ($\times 10^{-5}$)	+6	+2	−1	+4	−1

(*c*) Nickel. Plastic extension 14 %

plane reflecting	311	420	331	222
X-radiation used	Mn $K\alpha$	Cu $K\alpha$	Cu $K\alpha$	Mn $K\beta$
residual lattice strain ($\times 10^{-5}$)	−1	+3	−3	−11

(*d*) *Conclusions*

The results show that the residual lattice strains shown by different X-ray reflexions from a plastically extended polycrystalline metal wire vary both in magnitude and in sign. This is in qualitative agreement with the proposed explanation.

3. QUANTITATIVE INTERPRETATION

(*a*) *Introduction*

To establish that the observed residual lattice strains were of the order of magnitude to be expected, an attempt was made to calculate the values from the available knowledge of the process of plastic deformation in an aggregate. An exact treatment of the problem is too complicated, and consequently simplifying assumptions were made. The metal crystals were treated as elastically isotropic, and the normal 'mechanical' values of the elastic constants used. It was also assumed that, since the metals were all face-centred cubic and plastically deformed by tension, the deformation took place entirely by glide along the glide planes and in the glide directions (called later simply glide combinations) when the shear stress reached a critical value.

Further, it must be postulated which of the twelve glide combinations possible in the case of a face-centred cubic metal are operative in any given crystal in the aggregate. Cox & Sopwith (1937) assumed that each crystal behaved as if isolated from its neighbours, and that the operative glide combination was that in which the shear stress was greatest when the deforming tensile stress was parallel to the load applied to the aggregate.

Taylor (1938) pointed out that this would lead to the development of holes in the aggregate, and suggested that each crystal deforms so that its external shape changed in exactly the same way as that of the aggregate as a whole. Taylor showed that, to fulfil this condition, in general five motions in the glide combinations must take place, and he assumed that the operative combination was that requiring least work; with these assumptions he successfully derived the stress *v.* strain curve for an aggregate from that of a single crystal.

Calculations of the residual lattice strains to be expected were made with each of these two assumptions. Details are recorded here of those based on that of Taylor, since these gave the better agreement with experiment; only the results are given of the calculation based on the assumption that each grain was deforming as a single crystal.

(*b*) *Theoretical treatment*

Suppose we represent the polycrystalline aggregate by a model in which the stress P applied to the specimen is taken as acting on two rigid parallel plates which are connected by crystals of various orientations. If the crystals in the aggregate are randomly oriented, the cross-sectional area A of each crystal in the aggregate is the same. The model is represented in figure 2.

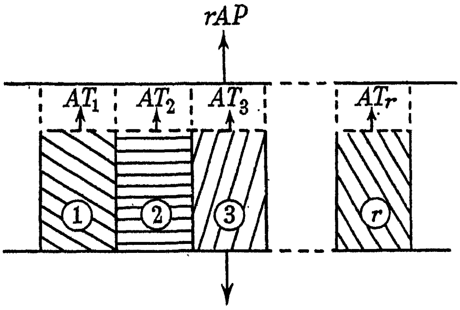


FIGURE 2. Model of polycrystalline aggregate. P is the stress applied to the aggregate. T_1, T_2, etc., are the stresses in the crystals of orientations 1, 2, etc.

Suppose P is increased from zero. Initially, all the crystals deform elastically and, since the crystals are assumed to be elastically isotropic, $T_1 = T_2 =$ etc. The yield point of the grains with the lowest yield tension, say grain 1, will be exceeded first, and the strains in these grains will then not be entirely elastic: $T_1 < T_2$, etc. After the aggregate has been given a certain strain, all the grains will have deformed plastically, since the limits of elastic strain for the grains with even the highest yield tension are small compared with the plastic strains. At this time, T_1, T_2, etc., will be determined by the orientations of the crystals, and $rAP = AT_1 + AT_2 + \dots + AT_r$.

If P is now removed, each crystal will contract elastically, and grains in which T was small, e.g. grain 1, will be driven into compression by grains in which T was large. The compressive stress in grain 1 may rise so high that plastic deformation would occur in the reverse direction in this grain, but a consideration of the magnitude of the stresses indicates that in a cubic metal this cannot occur often.

When equilibrium is established, the residual stress in any crystal n will be

$$T_n - T_{\text{mean}},$$

where T_n was the tensile stress in the crystal n when P was applied, T_{mean} was the average value of $T_1, T_2, \dots, T_n, \dots, T_r$ i.e. $T_{\text{mean}} = P$.

As pointed out in § 2 (c), the experimental arrangement was such that the residual lattice strain was measured in a direction perpendicular to the applied stress; thus the residual lattice strain in grain n is

$$-\frac{\nu}{E}(T_n - T_{\text{mean}}),$$

where ν is Poisson's ratio, and E Young's modulus.

T_{mean} is equal to the stress deforming the aggregate. T_n may be expressed in terms of the orientation and critical shearing stress (τ_c) of the crystal, and τ_c may in turn be expressed in terms of T_{mean}.

The critical shearing stress, τ_c, increases during the deformation ('work-hardening'), and, since some crystals undergo more shear than their neighbours, the amount of work-hardening is different for different crystals. Thus, after a given plastic deformation of the aggregate, τ_c is not the same for all crystals. But for the case of aluminium, an examination of the stress against strain curve in Taylor's paper for a single crystal, in conjunction with calculations taking into account the varying shear strains for different grains, indicates that after a longitudinal extension of each grain of 14 %, the maximum variation of τ_c is 15 %. The error involved in assuming τ_c to be a constant for all crystals in the plastically deformed aluminium aggregate is not large. The assumption is also made for the cases of copper and nickel.

We now require the relation between τ_c and T_1, T_2, etc. If the crystal in the aggregate behaves as if isolated from its neighbours, then

$$\tau_c = T \sin\chi \cos\lambda,$$

where χ is the angle between the applied stress and the most favourably oriented glide plane, and λ is the angle between the applied stress and the glide direction.

If the strain in each crystal is the same as that of the aggregate, then, during the plastic deformation, there must be transverse forces acting on each grain in addition to the stress T parallel to the direction of elongation. In Taylor's original analysis, it was not necessary to determine the stresses acting on any grain, because the work done by the force applied to the whole aggregate was equated to the work done against the shearing forces across the glide planes. In the present treatment, however, a knowledge of the magnitude of the stresses acting on each grain is necessary. But the precise determination of these stresses is complex, and consequently, for the purpose of this simple treatment which is only intended to give approximate values, the transverse stresses are neglected. This is the most serious source of error, and its possible effect is discussed in § 4.

The longitudinal stress acting on a crystal is taken as that which does work on the crystal equal to that performed against the shearing forces. For deformations which are small enough to keep the cross-sectional area and orientation of the crystal approximately constant, the stress parallel to the direction of extension, T, is given by

$$Tx = \tau_c \Sigma s,$$

where x = the longitudinal strain of the aggregate (and hence of the crystal), and, Σs = the arithmetical sum of the five shear strains required to produce the deformation. τ_c is assumed to be the same for all grains and x is the same in each case, consequently

$$T_1 = \frac{\tau_c}{x}(\Sigma s), \text{ etc.}$$

Since measurements on the diffraction lines were made so that the reflected X-ray beams observed lay perpendicular to the applied stress, the crystals investigated were oriented so that the direction of the reflecting normal was perpendicular to the applied stress, but they had any orientation about this normal. For the crystals contributing to an *hkl* diffraction line, the value of T required is

$$T_{hkl} = \frac{T_1 + T_2 + \ldots + T_n}{n},$$

where orientations $1, 2, \ldots, n$ are all those possible when *hkl* is fixed in direction:

$$T_{hkl} = \frac{\tau_c}{x}\left\{\frac{(\Sigma s)_1 + (\Sigma s)_2 + \ldots + (\Sigma s)_n}{n}\right\}$$

$$= \frac{\tau_c}{x}(\Sigma s)_{hkl}.$$

Similarly, $$T_{\text{mean}} = \frac{\tau_c}{x}\left\{\frac{(\Sigma s)_1 + (\Sigma s)_2 + \ldots + (\Sigma s)_n + \ldots + (\Sigma s)_r}{r}\right\}$$

$$= \frac{\tau_c}{x}(\Sigma s)_{\text{mean}},$$

where orientations $1, 2, \ldots, n \ldots, r$, are all those possible in the aggregate.

This last relation will be used to determine τ_c/x, T_{mean} being equal to the stress applied to the aggregate to produce the plastic extension. The value of τ_c/x applicable to the specimens examined is

$$\frac{T_{\text{mean}}}{(\Sigma s)_{\text{mean}}}.$$

Hence, the residual lattice strain shown by the *hkl* reflexion is

$$\left(\frac{\delta d}{d}\right)_{hkl} = \frac{\nu}{E} T_{\text{mean}}\left\{1 - \frac{(\Sigma s)_{hkl}}{(\Sigma s)_{\text{mean}}}\right\}.$$

Thus, if the residual lattice strains for various reflecting planes are plotted against $1 - \frac{(\Sigma s)_{hkl}}{(\Sigma s)_{\text{mean}}}$, the points should lie on a straight line through the origin with gradient $\nu T_{\text{mean}}/E$.

If one assumes the crystals in the aggregate to deform as if isolated from their neighbours, the corresponding relation is

$$\left(\frac{\delta d}{d}\right)_{hkl} = \frac{\nu}{E} T_{\text{mean}}\left\{1 - \frac{\left(\frac{1}{\sin\chi\cos\lambda}\right)_{hkl}}{\left(\frac{1}{\sin\chi\cos\lambda}\right)_{\text{mean}}}\right\}.$$

(c) *Results*

In order to calculate the numerical values of $(\Sigma s)_{hkl}$, etc., the data were arranged on a stereogram. The geometrical conditions determined by the experimental arrangement are that the reflecting normal (hkl) and the direction of the applied tension are fixed in space, and the crystal may have any orientation about the (hkl) normal. Figure 1 represents these conditions; the tension is applied in a direction perpendicular to the plane of the diagram.

On the stereogram it is convenient to consider the crystal as fixed in orientation; the locus of the applied tension is then a great circle about the (hkl) pole. The various orientations 1, 2, etc., were taken as points 5° apart on the locus when calculating $(\Sigma s)_{hkl}$; when calculating $(\Sigma s)_{\text{mean}}$, points were taken at 5° intervals over the smallest area of the stereogram from which the whole area can be built up by using the symmetry elements. Figure 3 shows the stereogram.

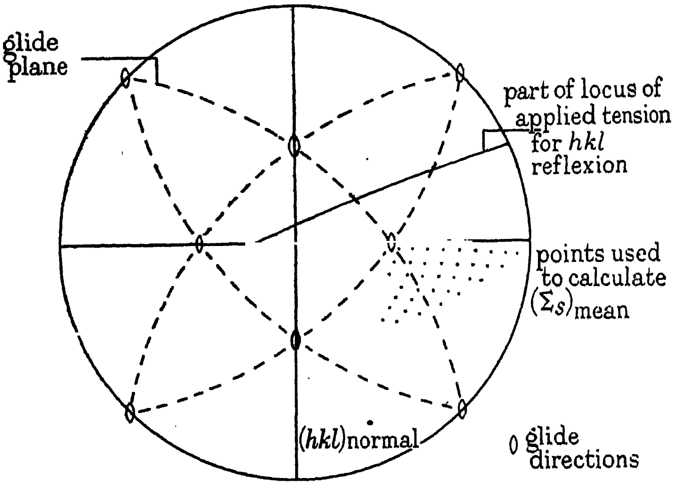

glide plane
part of locus of applied tension for hkl reflexion
points used to calculate (Σs)mean
(hkl) normal
0 glide directions

FIGURE 3. Experimental data arranged on stereogram.

The calculation of the values of Σs corresponding to $x = 0{\cdot}272$ has been performed by Taylor for a set of points on a stereogram. The value of x used is immaterial in this work, since relative, and not absolute, magnitudes of the various Σs values are required. The value of Σs for any point on the locus of the applied tension, or in the area, was obtained by interpolating between the values given by Taylor. A mean of the values of Σs on one locus gave $(\Sigma s)_{hkl}$, and in the area gave $(\Sigma s)_{\text{mean}}$. The results are given in table 2.

TABLE 2. VALUES OF Σ_s FOR FACE-CENTRED CUBIC METALS (FOR EXTENSION OF AGGREGATE $x = 0{\cdot}272$)

$(\Sigma s)_{\text{mean}} = 0{\cdot}836.$

plane reflecting	311	222	400	331	420	422	511 333
$(\Sigma s)_{hkl}$	0·814	0·907	0·805	0·863	0·789	0·864	0·811
$1-\dfrac{(\Sigma s)_{hkl}}{(\Sigma s)_{\text{mean}}}$ $(\times 10^{-2})$	+2·6	−8·5	+3·7	−3·2	+5·6	−3·3	+3·0

4. Comparison of experiment with theory

It must be emphasized that the calculations were of a very approximate nature, and, hence, appreciable differences between the experimental and calculated values were expected. It would be satisfactory if the points were such that a straight line could be drawn amongst them, a line of which the gradient was approximately that expected and which passed near the origin. Figures 4, 5 and 6 show the plotted points and the best straight line passing through them.

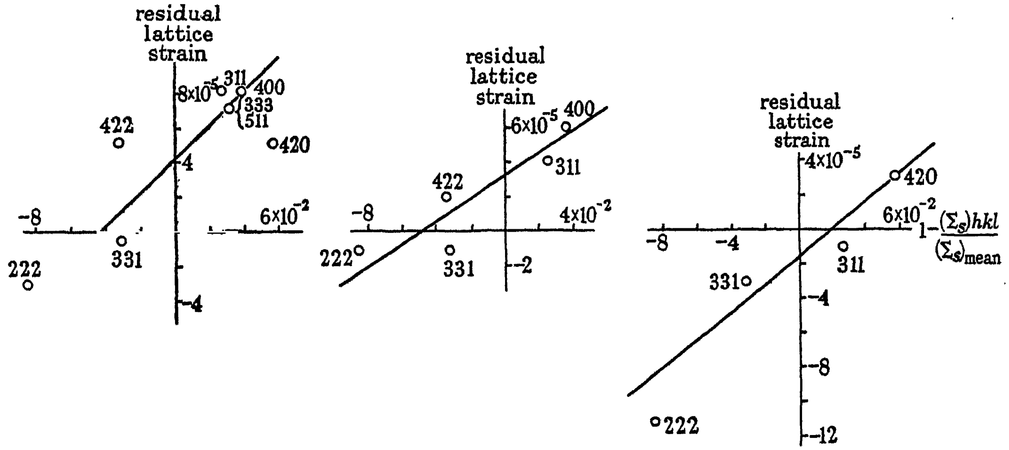


FIGURE 4. Aluminium. FIGURE 5. Copper. FIGURE 6. Nickel.

FIGURES 4 to 6. Relation between residual lattice strain and $1-(\Sigma s)_{hkl}/(\Sigma s)_{\text{mean}}$ for aluminium, copper and nickel respectively.

The principal features of each of these graphs are the linearity of the plotted points, the gradient of the best straight line through the points, and the intersection of this line and the axes. The effect of neglecting the transverse stresses required to produce the same strain in each crystal as in the aggregate will be that the plotted points will depart appreciably from the best straight line; the effect on the gradient and intercepts of the line is likely to be small, since these values depend on the positions of several points and averaging over several orientations makes the assumption more nearly true. It is convenient to separate further discussion of the three principal features.

(a) *Linearity of plotted points*

In general, the points depart considerably from the straight-line relationship, but in no case is the distribution of points on the graph completely random. The points indicate that there is a relation between the residual lattice strain and the quantity $1-(\Sigma s)_{hkl}/(\Sigma s)_{\text{mean}}$, approximately that of a straight line.

If the residual lattice strains are plotted against the function calculated with the assumption that each crystal in the aggregate behaves as if isolated, the points show slightly greater displacements from the best straight line.

(b) *Intersection of the straight line and the axes*

In the case of the graph for aluminium and for copper, it is apparent that the straight line cuts the residual strain axis at a positive value which is larger than the limits of experimental error and of possible errors due to assumptions made in the theory. In the case of nickel, the intercept is probably within the limits of error. The line through the points showing the relation between the residual lattice strain and the function calculated on the assumption of isolated crystals makes approximately the same intercept on the axis. This effect would arise if there were, in the grains examined, a macroscopic compressive stress superimposed on the calculated Heyn stresses. To balance this stress there would have to be a macroscopic tensile stress component in some part of the specimen not investigated.

A possible explanation of the positive intercept may be that the grains in which the residual lattice strains are being measured all lie near the surface, whereas the theory implicitly assumes that all grains in the specimen are surrounded by constraining neighbours. Crystals with a free surface would have a lower yield tension than would crystals of the same orientation in the interior, and hence, after the plastic extension of the aggregate, a compressive stress would be expected on the surface superimposed on the microscopic Heyn stresses. This idea is supported by some work at present in progress which indicates that the relative size of the specimen and the irradiated volume has an effect on the magnitude of the intercept on the residual strain axis.

It is also possible to explain this superimposed compressive stress in terms of arrested dislocations, using the ideas put forward by Bragg *et al.* (see p. 557), but the later experimental observation that the magnitude of the compressive stress varies with the specimen size does not appear to support this type of explanation.

It is hoped that further work will elucidate the nature and cause of this compressive stress superimposed on the Heyn stresses.

(c) *Gradients of the graphs*

The gradients of the best straight lines passing through the plotted points were measured, and the value $\nu T_{\text{mean}}/E$ calculated for each metal. Since the two possible assumptions on the interaction of the grains lead to different results, a comparison of the experimental gradient with the calculated value is shown in table 3 for each case.

TABLE 3. COMPARISON OF GRAPH GRADIENTS WITH THEORETICAL VALUE

	gradient measured on graph		
metal	crystal strained as aggregate ($\times 10^{-3}$)	crystal deforming as if isolated ($\times 10^{-3}$)	calculated $\nu T_{\text{mean}}/E$ ($\times 10^{-3}$)
aluminium	$0{\cdot}9_6$	$1{\cdot}4_1$	0·83
copper	$0{\cdot}6_2$	$1{\cdot}2_8$	0·71
nickel	$0{\cdot}7_9$	1·8	0·76

Table 3 shows that the theory based on Taylor's assumption that the strain in the grains is the same as that of the aggregate gives an agreement with experiment within the limits of error. The assumption of isolated grains does not give a satisfactory agreement.

5. CONCLUSIONS

Mean residual lattice strains, determined from measurements made on several diffraction lines from plastically extended aluminium, copper and nickel, have values which vary both in magnitude and in sign. This is qualitatively in agreement with the facts that a complex intergranular stress system exists in the deformed aggregate arising from the variation of yield tension with the orientation of the grains, and that in the diffraction process the X-rays select grains with a limited range of orientation. In view of the number of assumptions made in the quantitative treatment, one can say that the agreement between the experimental mean residual lattice strains and those calculated is satisfactory.

I should like to express my thanks to all my friends for their help and advice during the course of the work, and particularly to Dr E. Orowan, F.R.S., for much help with the preparation of this paper. I am indebted to Miss E. M. Chapman, of R.A.E., for the measurement of the residual lattice strains for nickel quoted in table 1 (*c*).

This paper is published with the permission of the Chief Scientist, Ministry of Supply.

REFERENCES

Bollenrath, F., Hauk, V. & Osswald, E. 1939 *Z. Ver. dtsch. Ing.* **83**, 129.
Cox, H. L. & Sopwith, D. G. 1937 *Proc. Phys. Soc.* **49**, 134.
Davidenkov, N. N. & Timofeeva, M. N. 1946 *J. Tech. Phys. U.S.S.R.* **16**, 283.
Greenough, G. B. 1947 *Nature*, **160**, 258.
Greenough, G. B. 1949 *Metal Treatm.* **16** (in the Press).
Smith, S. L. & Wood, W. A. 1941 *Proc. Roy. Soc.* A, **178**, 93.
Smith, S. L. & Wood, W. A. 1944 *Proc. Roy. Soc.* A, **182**, 404.
Taylor, G. I. 1938 *J. Inst. Met.* **62**, 307.

Electroviscosity
III. Sedimentation phenomena in ionic liquids

By G. A. H. Elton, *Chemistry Department, Battersea Polytechnic*

(*Communicated by J. Kenyon, F.R.S.—Received* 29 *January* 1949)

It is demonstrated that the rate of sedimentation of a system of particles in an ionic liquid may vary markedly with the ionic concentration, due to variation in the electroviscous forces which resist shear in the double layer of ions associated with each particle. Equations governing the rate of sedimentation are derived, and used to calculate values for the electrokinetic potentials of carborundum in potassium chloride solutions, from data on sedimentation velocity. These values are compared with those previously obtained from data on electro-endosmosis.

1. Introduction

It is well known that when an electrical double layer at an interface is sheared, a potential gradient is set up in the plane of shear. It has been shown (Elton 1948 *a*, *b*) that in the case of an ionic liquid flowing through a narrow channel, the electrical resistance to shear of the double layer may be of the same order of magnitude as the mechanical (viscous) resistance, with the result that the liquid has an abnormally large apparent viscosity. It is proposed to demonstrate in this paper that a similar effect becomes noticeable in the sedimentation of particles in suspension in an ionic liquid, and that the rate of sedimentation of a particle in a given liquid may provide useful information concerning the electrokinetic potential at the particle/liquid interface.

2. Sedimentation potentials

Consider a system of spherical particles of radius a and density ρ_2 sedimenting in an infinite volume of a solution of a uni-univalent electrolyte of ionic concentration n. Let this solution be of density ρ_1, and have an electrokinetic potential ζ at the particle/liquid interface. Let the particles be of concentration N per unit volume, being separated by distances such that each may be considered as a separate unit, electrically and mechanically. Let the specific conductivity of the suspension as a whole be κ. This will be very nearly equal to the bulk conductivity of the solution. As the particles fall through the liquid, a sedimentation potential gradient will be set up, given by

$$E = \frac{4\pi a^2 \sigma u N}{\kappa}, \tag{1}$$

where σ is the charge per unit area on a particle, and u is the velocity of sedimentation.

The charge carried by the particle is equal to the net charge in the associated double layer, and provided that the radius of the particle is much larger than the effective thickness of the double layer, we may write

$$\sigma = -\int_0^\infty \rho_x dx, \tag{2}$$

where ρ_x is the charge density at a plane distance x from the particle surface, and is given by the Poisson equation for a plane surface

$$\frac{d^2\psi}{dx^2} = -\frac{4\pi\rho_x}{\epsilon}, \tag{3}$$

where ψ is the potential at distance x from the surface, and ϵ is the dielectric constant at that point, usually taken as equal to the bulk dielectric constant of the liquid.

From equations (2) and (3)

$$\sigma = \frac{\epsilon}{4\pi}\int_0^\infty \frac{d^2\psi}{dx^2}dx, \tag{4}$$

that is,

$$\sigma = \frac{\epsilon}{4\pi}\left[\frac{d\psi}{dx}\right]_0^\infty. \tag{5}$$

Since $d\psi/dx = 0$ when $x = \infty$, this becomes

$$\sigma = -\frac{\epsilon}{4\pi}\left(\frac{d\psi}{dx}\right)_{x=0}. \tag{6}$$

Verwey & Overbeek (1948) show that this leads to

$$\sigma = \left(\frac{2n\epsilon kT}{\pi}\right)^{\frac{1}{2}}\sinh\left(\frac{e\psi_0}{2kT}\right), \tag{7}$$

where $\boldsymbol{k}$ is the Boltzmann constant, T the absolute temperature, e the charge on the electron and $\psi_0 = \zeta$. From equations (1) and (7), and setting $u = \dfrac{2a^2(\rho_2-\rho_1)g}{9\eta_a}$, where η_a is the total apparent viscosity, we obtain the expression for the sedimentation potential

$$E = \frac{8\pi a^4(\rho_2-\rho_1)gN}{9\eta_a\kappa}\left(\frac{2n\epsilon kT}{\pi}\right)^{\frac{1}{2}}\sinh\left(\frac{e\zeta}{2kT}\right). \tag{8}$$

When $\dfrac{e\zeta}{2kT} \ll 1$, that is, when $\zeta \ll 50\,\mathrm{mV}$, we may write $\sinh\left(\dfrac{e\zeta}{2kT}\right) = \dfrac{e\zeta}{2kT}$, when equation (8) reduces to

$$E = \frac{2a^4(\rho_2-\rho_1)gN\epsilon\zeta}{9\eta_a\kappa\lambda}, \tag{9}$$

where $\lambda = \left(\dfrac{\epsilon kT}{8\pi ne^2}\right)^{\frac{1}{2}}$, the Debye-Hückel 'double-layer thickness'.

Although this approximation is only strictly valid for very small electrokinetic potentials, the work of Verwey & Overbeek indicates that it may adequately represent experimental data at rather higher values.

In cases where the double-layer thickness is not negligible compared with the radius of the particle, Verwey & Overbeek show that the charge per unit area on the particle is for small values of ζ approximately

$$\sigma = \frac{\epsilon\zeta}{4\pi a}\left(1+\frac{a}{\lambda}\right). \tag{10}$$

3. ELECTROVISCOUS RESISTANCE TO SEDIMENTATION

The rate of sedimentation of a particle in the suspension as postulated will be smaller than it would be in a non-ionic liquid having the same density and viscosity, since it is subjected not only to viscous retarding forces, but also to an electrical retarding force exerted on its associated charge by the potential through which it falls. At equilibrium, that is, for uniform rate of sedimentation,

$$6\pi\eta_a au = \tfrac{4}{3}\pi a^3(\rho_2-\rho_1)\,g = 6\pi\eta au + E\,4\pi a^2\sigma, \tag{11}$$

where η is the viscosity of the suspension in the absence of electrical retarding forces, and η_a is the total apparent viscosity.

Hence, substituting for E, σ and u, we have

$$\eta_a = \eta + \frac{16n\epsilon kTa^3N}{3\pi\kappa}\sinh^2\left(\frac{e\zeta}{2kT}\right). \tag{12}$$

For small values of ζ this reduces to

$$\eta_a = \eta + \frac{a^3\epsilon^2\zeta^2N}{6\pi\lambda^2\kappa}. \tag{13}$$

It is shown experimentally in the following section that the difference between η and η_a may be considerable. The effect of increasing the salt concentration of the solution will be to decrease ζ, and at high concentrations, when ζ tends to zero, η_a will tend to η.

In cases where λ is not negligible compared with a, the charge per unit area on the particle is given approximately by equation (10), and equation (13) becomes

$$\eta_a = \eta + \frac{a\epsilon^2\zeta^2N(1+a/\lambda)^2}{6\pi\kappa}. \tag{14}$$

4. EXPERIMENTAL

In order to test experimentally the equations derived above, it was proposed to make observations on the rates of sedimentation of two available specimens of carborundum. Fairbrother & Mastin (1924, 1925) have studied electro-osmosis of various electrolytes in diaphragms of carborundum, and used the Helmholtz-Smoluchowski equation

$$v = \frac{\zeta\epsilon E}{4\pi\eta}, \tag{15}$$

where v is the velocity of electro-endosmosis under a potential gradient E to calculate electrokinetic potentials from their data. These values for ζ are thus available for comparison with values of ζ which may be calculated from equation (12) if the rate of sedimentation of particles in suspension in a liquid of given concentration is known, and compared with the rate of sedimentation in the absence of electroviscous forces, that is, as the concentration tends to infinity.

The main difficulty to be expected in calculating the electrokinetic potential in this way lies in the fact that the particles are not spherical. Since the irregularly shaped particles would have a larger surface area than spheres of the same effective radius (as determined by sedimentation velocity), the charge transported will be greater, and the value of ζ calculated on the assumption that the particles are spheres will be larger than the true value. However, if the equations derived above represent correctly the conditions governing sedimentation, values for the electrokinetic potentials at least of the right order should be obtained.

The two carborundum specimens used were separated by successive sedimentation from 600- and 800-mesh material supplied by the Carborundum Company, Ltd., and microscopic observation indicated that they were of uniform particle size. The specimens were treated with aqua regia, and washed many times with distilled water, followed by several washings with conductivity water. AnalaR potassium chloride recrystallized from conductivity water was used in the preparation of the solutions. The specific conductivity of the water used was about $0{\cdot}6 \times 10^{-6}$ ohm^{-1} cm.$^{-1}$.

The apparatus in which sedimentation was carried out consisted simply of a measuring cylinder of length 30 cm. and diameter 3·5 cm., fitted with a ground-glass stopper. Into this was introduced 200 ml. of the solution to be used, containing 1×10^{-3} g./ml. of carborundum in suspension. The tube was clamped vertically in a thermostat maintained at $25{\cdot}00 \pm 0{\cdot}005^\circ$ C until temperature equilibrium had been reached, then removed for a few moments and thoroughly shaken to bring the carborundum into a uniform state of suspension. The tube was replaced vertically in the thermostat, and the sedimentation observed. It was seen after a short time that there was a sharp division between the sedimenting suspension and the clear liquid, giving a meniscus that was more or less concave, due probably to wall effects and to the extra time taken for the particles to detach themselves initially from the edge of the air/solution interface. The bottom of the meniscus, in the middle of the tube, was observed, and the time taken to fall between various fixed marks on the tube determined. A graph of distance below the surface of the solution against time was plotted for at least two determinations on each system, a straight line being obtained after a few minutes. The slope of this line was determined by the method of least squares from each set of data, results being reproducible to about 2 %.

It was found that as the salt concentration was increased from very small values, the rate of sedimentation of a given suspension increased markedly. When the concentration had reached a value of about 1×10^{-3} N, the rate of sedimentation was only slightly increased by further additions of salt. Data for the two suspensions are given in table 1, together with an extrapolated value for R_∞, the limiting rate of sedimentation in concentrated solutions. Values for the electrokinetic potentials of the carborundum in the various solutions were then calculated both from equation (12) and from equation (13) in each case, making the simplifying assumption that the particles were spherical. The results of these calculations are given in table 2, together with the values obtained at the same concentration by interpolation from the data of Fairbrother & Mastin. The final column of table 2 shows the ratio between the mean values of ζ as determined by sedimentation, using the exact equation (12), and ζ as determined by electro-endosmosis. It is seen that the values calculated from the data using equation (13) are very high, since the approximation

$$\sinh\left(\frac{e\zeta}{2kT}\right) = \frac{e\zeta}{2kT}$$

is not in these cases even approximately valid. Equation (12) leads to values of ζ which are, as expected, higher than those obtained from electro-endosmosis data, but in an almost constant ratio. These experiments thus provide a semi-quantitative verification of the theoretical equations.

TABLE 1

KCl normality	batch I (rate mm./min.)	batch II (rate mm./min.)
1×10^{-5}	$0{\cdot}51_8$	$0{\cdot}12_5$
2×10^{-5}	$0{\cdot}66_1$	$0{\cdot}15_8$
5×10^{-5}	$0{\cdot}91_2$	$0{\cdot}21_0$
1×10^{-4}	$1{\cdot}07_0$	$0{\cdot}26_6$
2×10^{-4}	$1{\cdot}20_8$	$0{\cdot}30_0$
3×10^{-4}	$1{\cdot}30_9$	$0{\cdot}31_1$
5×10^{-4}	$1{\cdot}36_1$	$0{\cdot}33_3$
1×10^{-3}	$1{\cdot}41_5$	$0{\cdot}34_1$
5×10^{-3}	$1{\cdot}44_5$	$0{\cdot}34_9$
1×10^{-2}	$1{\cdot}45_6$	$0{\cdot}36_0$
R_∞ (ext.)	1·48	0·36

TABLE 2

	batch I		batch II		mean $\zeta_{\text{sed.}}$ from eqn. (12) (mV)	$\zeta_{\text{end.}}$ (mV) Fair-brother & Mastin	$\zeta_{\text{sed.}}/\zeta_{\text{end.}}$
KCl normality	ζ from eqn. (13) (mV)	ζ from eqn. (12) (mV)	ζ from eqn. (13) (mV)	ζ from eqn. (12) (mV)			
1×10^{-5}	2713	265	2730	265	265	—	—
2×10^{-5}	2223	254	2244	255	255	73	3·49
5×10^{-5}	1521	236	1623	240	238	68	3·50
1×10^{-4}	1230	224	1181	222	223	66	3·38
2×10^{-4}	939	210	893	210	210	63	3·33
3×10^{-4}	718	196	788	201	199	60	3·32
5×10^{-4}	582	187	555	183	185	56	3·30
1×10^{-3}	424	170	458	174	172	50	3·44

REFERENCES

Elton, G. A. H. 1948*a* *Proc. Roy. Soc.* A, **194**, 259.
Elton, G. A. H. 1948*b* *Proc. Roy. Soc.* A, **194**, 275.
Fairbrother, F. & Mastin, H. 1924 *J. Chem. Soc.* **125**, 2319.
Fairbrother, F. & Mastin, H. 1925 *J. Chem. Soc.* **127**, 322.
Verwey, E. J. W. & Overbeek, J. Th. G. 1948 *Theory of the stability of lyophobic colloids.* London: Elsevier Publishing Company, Inc.

INDEX TO VOLUME 197 (A)

Acetaldehyde hydrate, catalyzed dehydration (Bell & Higginson), 141.
Aluminium-cobalt-copper alloys, constitution (Pfeil & Raynor), 321.
Andrade, E. N. da C. The presentation of scientific information, 1.
Application of semi-conductivity measurements in the study of catalysis on copper oxide (Gray), 314.
Arnold, R. N. & Warburton, G. B. Flexural vibrations of the walls of thin cylindrical shells having freely supported ends, 238.
Atomic parameters, refinement of (Booth), 336.
Axford, D. W. E. Studies in polymerization. IV. The velocity constants in polymerization of *p*-methoxystyrene, 374.

Badger, E. H. M., Griffith, R. H. & Newling, W. B. S. The catalytic decomposition of simple heterocyclic compounds. I. Preparation and properties of molybdenum sulphide and nickel sulphide catalysts, 184.
Bamford, C. H. & Dewar, M. J. S. Studies in polymerization. III. The polymerization of methyl methacrylate, 356.
Bell, R. P. & Higginson, W. C. E. The catalyzed dehydration of acetaldehyde hydrate, and the effect of structure on the velocity of protolytic reactions, 141.
Binnie, A. M. The passage of a perfect fluid through a critical cross-section or 'throat', 545.
Bond lengths in certain molecules (Hume-Rothery), 17.
Booth, A. D. The refinement of atomic parameters by the technique known in X-ray crystallography as 'the method of steepest descents', 336.
Boys, S. F. & Corner, J. The structure of the reaction zone in a flame, 90.
Bullard, E. C. Magnetic field within the earth, 433.
Burhop, E. H. S. & Yadav, H. N. The scattering of high-energy neutrons by protons and the law of interaction between nucleons, 505.

Catalysis on copper oxide, study of (Gray), 314.
Catalytic decomposition of simple heterocyclic compounds. I. Preparation and properties of molybdenum sulphide and nickel sulphide catalysts (Badger, Griffith & Newling), 184.
Catalytic decomposition of simple heterocyclic compounds. II. Reaction kinetics and mechanism (Griffith, Marsh & Newling), 194.
Catalyzed dehydration of acetaldehyde hydrate, and the effect of structure on the velocity of protolytic reactions (Bell & Higginson), 141.
Charge distribution and energy levels of trapped electrons in ionic solids (Simpson), 269.
Constitution of the aluminium-rich aluminium-cobalt-copper alloys, with special reference to the role of transitional metals in alloy formation (Pfeil & Raynor), 321.
Cooper, R. I. B. Recent advances in the technique of submarine gravity surveying, 523.
Cope, W. F. The hydrodynamical theory of film lubrication, 201.
Corner, J. *See* Boys & Corner.
Cox, D. R. Theory of drafting of wool slivers. I, 28.
Crystal growth from solution (Humphreys-Owen), 218.
Cybulski, W. B., Payman, W. & Woodhead, D. W. Explosion waves and shock waves. VII. The velocity of detonation in cast T.N.T., 51.

Davies, R. M. The determination of static and dynamic yield stresses using a steel ball, 416.
Determination of static and dynamic yield stresses using a steel ball (Davies), 416.

Dewar, M. J. S. *See* Bamford & Dewar.
Dislocations as a cause of mechanical damping in metals (Eshelby), 396.

Electrodeless discharges, starting potentials (Gill & von Engel), 107.
Electroviscosity. III. Sedimentation phenomena in ionic liquids (Elton), 568.
Elton, G. A. H. Electroviscosity. III. Sedimentation phenomena in ionic liquids, 568.
Eshelby, J. D. Dislocations as a cause of mechanical damping in metals, 396.
Evaporation and surface structure of liquids (Wyllie), 383.
Explosion waves and shock waves. VII. The velocity of detonation in cast T.N.T. (Cybulski, Payman & Woodhead), 51.

Film lubrication, hydrodynamical theory (Cope), 201.
Flexural vibrations of the walls of thin cylindrical shells having freely supported ends (Arnold & Warburton), 238.

Gabor, D. Microscopy by reconstructed wave-fronts, 454.
Garner, W. E., Gray, T. J. & Stone, F. S. The oxidation of copper and the reactions of hydrogen and carbon monoxide with copper oxide, 294.
Gill, E. W. B. & von Engel, A. Starting potentials of electrodeless discharges, 107.
Goody, R. M. The thermal equilibrium at the tropopause and the temperature of the lower stratosphere, 487.
Gravity surveying (submarine) technique (Cooper), 523.
Gray, T. J. *See* Garner, Gray & Stone.
Gray, T. J. The application of semi-conductivity measurements in the study of catalysis on copper oxide, 314.
Green, H. S. On the self-energies and cross-sections of orthodox quantum mechanics, 73.
Greenough, G. B. Residual lattice strains in plastically deformed polycrystalline metal aggregates, 556.
Griffith, R. H. *See* Badger, Griffith & Newling.
Griffith, R. H., Marsh, J. D. F. & Newling, W. B. S. The catalytic decomposition of simple heterocyclic compounds. II. Reaction kinetics and mechanism, 194.

Heterocyclic compounds, catalytic decomposition. I. (Badger, Griffith & Newling), 184.
Heterocyclic compounds, catalytic decomposition. II. (Griffith, Marsh & Newling), 194.
Higginson, W. C. E. *See* Bell & Higginson.
Hume-Rothery, W. On the bond lengths and interatomic distances in certain molecules and crystals, 17.
Humphreys-Owen, S. P. F. Crystal growth from solution, 218.
Hydrodynamical theory of film lubrication (Cope), 201.

Integral equations (Kreisel), 160.
Interatomic distances in certain molecules (Hume-Rothery), 17.

Kreisel, G. Some remarks on integral equations with kernels: $L(\xi_1-x_1, \ldots, \xi_n-x_n; \alpha)$, 160.

Magnetic field within the earth (Bullard), 433.
Marsh, J. D. F. *See* Griffith, Marsh & Newling.
Microscopy by reconstructed wave-fronts (Gabor), 454.
Momentum and energy diffusion in the turbulent wake of a cylinder (Townsend), 124.

Newling, W. B. S. *See* Badger, Griffith & Newling, *also* Griffith, Marsh & Newling.
Nucleons, law of interaction (Burhop & Yadav), 505.
Neutrons, scattering by protons (Burhop & Yadav), 505.

On the bond lengths and interatomic distances in certain molecules and crystals (Hume-Rothery), 17.
On the self-energies and cross-sections of orthodox quantum mechanics (Green), 73.
Oxidation of copper and the reactions of hydrogen and carbon monoxide with copper oxide (Garner, Gray & Stone), 294.

Passage of a perfect fluid through a critical cross-section or 'throat' (Binnie), 545.
Payman, W. *See* Cybulski, Payman & Woodhead.
Pfeil, P. C. L. & Raynor, G. V. The constitution of the aluminium-rich aluminium-cobalt-copper alloys, with special reference to the role of transitional metals in alloy formation, 321.
Polymerization studies. III (Bamford & Dewar), 356.
Polymerization studies. IV (Axford), 374.
Presentation of scientific information (Andrade), 1.
Protocatechuic acid crystals, spontaneous deformation (Wood), 283.

Quadratic form x^2-7y^2 (Varnavides), 256.
Quantum mechanics, self-energies and cross-sections (Green), 73.

Raynor, G. V. *See* Pfeil & Raynor.
Reactions zone in a flame, structure (Boys & Corner), 90.
Recent advances in the technique of submarine gravity surveying (Cooper), 523.
Refinement of atomic parameters by the technique known in X-ray crystallography as 'the method of steepest descents' (Booth), 336.
Residual lattice strains in plastically deformed polycrystalline metal aggregates (Greenough), 556.

Scattering of high-energy neutrons by protons and the law of interaction between nucleons (Burhop & Massey), 505.
Simpson, J. H. Charge distribution and energy levels of trapped electrons in ionic solids, 269.
Some remarks on integral equations with kernels: $L(\xi_1-x_1, \ldots, \xi_n-x_n; \alpha)$ (Kreisel), 160.
Spontaneous deformation of protocatechuic acid crystals (Wood), 283.
Starting potentials of electrodeless discharges (Gill & von Engel), 107.
Stone, F. S. *See* Garner, Gray & Stone.
Stratosphere, temperature of the lower (Goody), 487.
Structure of the reaction zone in a flame (Boys & Corner), 90.
Studies in polymerization. III. The polymerization of methyl methacrylate (Bamford & Dewar), 356.
Studies in polymerization. IV. The velocity constants in the polymerization of *p*-methoxystyrene (Axford), 374.

Theory of drafting of wool slivers. I (Cox), 28.
Thermal equilibrium at the tropopause and the temperature of the lower stratosphere (Goody), 487.
T.N.T., velocity of detonation (Cybulski, Payman & Woodhead), 51.
Townsend, A. A. Momentum and energy diffusion in the turbulent wake of a cylinder, 124.
Tropopause, thermal equilibrium (Goody), 487.
Turbulent wake of a cylinder (Townsend), 124.

Varnavides, P. On the quadratic form x^2-7y^2, 256.
von Engel, A. *See* Gill & von Engel.

Warburton, G. B. *See* Arnold & Warburton.
Wood, R. W. Spontaneous deformation of protocatechuic acid crystals, 283.
Woodhead, D. W. *See* Cybulski, Payman & Woodhead.
Wool slivers, theory of drafting (Cox), 28.
Wyllie, G. Evaporation and surface structure of liquids, 383.

Yadav, H. N. *See* Burhop & Yadav.

Yield stresses, determination (Davies), 416.

END OF THE ONE HUNDRED AND NINETY-SEVENTH VOLUME (SERIES A)

I.A.R.I. 75

INDIAN AGRICULTURAL RESEARCH
INSTITUTE LIBRARY, NEW DELHI.

Date of Issue	Date of Issue	Date of Issue
9.4.57		
20.9.55		

GIPNLK—H-40 I.A.R I.—29-4-55—15,000